

Structural model and properties of the AdoMetDC domain of  
the bifunctional *Plasmodium falciparum* S-adenosylmethionine  
decarboxylase/Ornithine decarboxylase

by

Gordon Andreas Wells

Submitted in partial fulfilment of the requirements for the degree *Magister Scientiae*

in the Faculty of Natural and Agricultural Sciences

Department of Biochemistry

University of Pretoria

Pretoria

January 2004

## Acknowledgements

- My supervisor Prof. A. I. Louw, and cosupervisor Dr Fourie Joubert of the University of Pretoria Biochemistry Department for enabling me to pursue this project and enter the world of structural biology.
- My cosupervisor Dr Lyn-Marie Birkholtz for always being prepared to give much needed advice and criticism.
- Prof. Walter of The Bernhard Nocht Institute (BNI) for Tropical Medicine, Hamburg, Germany for giving me the opportunity to visit his laboratory. This enabled me to perform the experiments described herein and allowed me to gain invaluable experience.
- My fellow students of the University of Pretoria and the students of the BNI for helping a self-confessed and incurable computer geek to get his hands wet doing real Biochemistry.
- Prof S. Ealick (Cornell University, USA) for providing the crystal structure of the potato enzyme prior to publication.
- The National Research Foundation (NRF) of South Africa and the University of Pretoria for awarding me the bursaries which enabled me to continue my studies.
- No Microsoft® employees were directly harmed in the making of this production: I am indebted to the Open Source Community for providing much of the software that has enabled me to do my work over the last three years<sup>1</sup>.
- My family for their support and patience. Specifically, my parents for making the invaluable purchase of a 486 DX2 50 Mhz computer 9 years ago.

---

<sup>1</sup> This manuscript and almost all figures were prepared using Open Source Software. Typesetting was done in L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>, using version 1.3 of the LyX editor.

# Contents

<b>Acknowledgements</b> . . . . .	i
<b>List of Figures</b> . . . . .	v
<b>List of Tables</b> . . . . .	vii
<b>Typographical conventions</b> . . . . .	viii
<b>List of Abbreviations</b> . . . . .	ix
<b>List of Computer Related Terms</b> . . . . .	xi
<b>Chapter 1. Introduction</b> . . . . .	1
1.1. The need for new anti-malarials . . . . .	1
1.2. Polyamines . . . . .	5
1.2.1. Functions of polyamines . . . . .	5
1.2.2. Polyamine metabolism . . . . .	6
1.2.3. Polyamines in malaria . . . . .	8
1.2.4. Polyamines as a drug target . . . . .	8
1.3. Properties of S-adenosylmethionine decarboxylase (AdoMetDC) . . . . .	9
1.3.1. AdoMetDC requires pyruvoyl . . . . .	9
1.3.2. Enzymatic mechanism . . . . .	12
1.3.2.1. Effects of putrescine on AdoMetDC . . . . .	12
1.3.3. Structure of AdoMetDC . . . . .	13
1.3.3.1. The AdoMetDC fold . . . . .	13
1.3.3.2. AdoMetDC structure, processing and enzyme activity . . . . .	14
1.3.3.3. AdoMetDC structure and putrescine stimulation . . . . .	16
1.3.4. Malarial AdoMetDC . . . . .	17
1.4. Aims . . . . .	18
<b>Chapter 2. Structural modelling of <i>P. falciparum</i> AdoMetDC</b> . . . . .	19
2.1. Introduction . . . . .	19
2.1.1. The need for Bioinformatics . . . . .	19
2.1.2. Computational protein modelling . . . . .	20
2.2. Methods . . . . .	22
2.2.1. Identification of other <i>Plasmodium</i> sequences . . . . .	22
2.2.2. Multiple alignment . . . . .	22

2.2.3.	Secondary structure prediction . . . . .	23
2.2.4.	Homology modelling . . . . .	23
2.3.	Results . . . . .	25
2.3.1.	Identification of other <i>Plasmodium</i> sequences . . . . .	25
2.3.2.	Alignment and motif identification . . . . .	26
2.3.3.	Secondary structure prediction . . . . .	27
2.3.3.1.	Inserts . . . . .	27
2.3.4.	Homology modelling . . . . .	28
2.3.4.1.	Overall model characteristics . . . . .	28
2.3.4.2.	Active site residues . . . . .	31
2.3.4.3.	Active site shape . . . . .	35
2.3.4.4.	Structure of insert 1 . . . . .	36
2.4.	Discussion . . . . .	37
2.4.1.	Identification of other <i>Plasmodium</i> sequences . . . . .	37
2.4.2.	Sequence properties of the <i>Plasmodium</i> AdoMetDC/ODC . . . . .	38
2.4.2.1.	Conservation of secondary structural elements . . . . .	38
2.4.2.2.	<i>Plasmodium</i> -specific inserts . . . . .	39
2.4.3.	Homology modelling . . . . .	41
2.4.3.1.	Overall model characteristics . . . . .	41
2.4.3.2.	Active site composition . . . . .	42
2.4.3.3.	Active site shape . . . . .	44
2.4.3.4.	Structure of insert 1 . . . . .	44
<b>Chapter 3. Model guided mutational analysis of malarial AdoMetDC . . . . .</b>		<b>46</b>
3.1.	Introduction . . . . .	46
3.2.	Methods . . . . .	48
3.2.1.	<i>In silico</i> putrescine docking . . . . .	48
3.2.2.	Construction of putrescine-like mutants . . . . .	48
3.2.2.1.	Mutagenesis of wild-type bifunctional AdoMetDC/ODC plasmid construct . . . . .	48
3.2.2.2.	Sequencing of putrescine-like mutants . . . . .	49
3.2.3.	Recombinant expression . . . . .	50
3.2.4.	Enzyme Assays . . . . .	51
3.3.	Results . . . . .	51
3.3.1.	Putrescine docking . . . . .	51
3.3.1.1.	Comparison with human and model residues . . . . .	51
3.3.1.2.	Putrescine docking . . . . .	52
3.3.2.	Mutagenesis of wild-type bifunctional AdoMetDC/ODC . . . . .	53
3.3.3.	Expression of mutant AdoMetDC/ODC . . . . .	54
3.4.	Discussion . . . . .	54
3.4.1.	Putrescine docking . . . . .	54



3.4.2. Mutagenesis . . . . .	56
<b>Chapter 4. Model guided inhibitor screening of malarial AdoMetDC . . . . .</b>	<b>58</b>
4.1. Introduction . . . . .	58
4.1.1. <i>In silico</i> ligand docking . . . . .	58
4.2. Methods . . . . .	60
4.2.1. <i>In silico</i> inhibitor screening . . . . .	60
4.2.2. Test compound solutions . . . . .	60
4.2.3. Assays . . . . .	61
4.3. Results . . . . .	61
4.3.1. <i>In silico</i> inhibitor screening . . . . .	61
4.3.2. Solubility of potential inhibitors . . . . .	61
4.3.3. Inhibition of AdoMetDC . . . . .	63
4.4. Discussion . . . . .	65
<b>Chapter 5. Concluding Discussion . . . . .</b>	<b>67</b>
<b>Summary . . . . .</b>	<b>72</b>
<b>Opsomming . . . . .</b>	<b>73</b>
<b>Bibliography . . . . .</b>	<b>74</b>
<b>Appendix A. Supplementary data for chapter 2 . . . . .</b>	<b>85</b>
CLUSTALX protein colouring: . . . . .	85
Swissprot accession numbers for multiple sequence alignment . . . . .	85
<b>Appendix B. Supplementary data for chapter 4 . . . . .</b>	<b>91</b>

## List of Figures

1.1.	The life cycle of <i>P. falciparum</i> . . . . .	2
1.2.	Current global status of malaria resistance . . . . .	3
1.3.	The main polyamines . . . . .	5
1.4.	The generic polyamine pathway . . . . .	7
1.5.	Formation of the AdoMetDC pyruvoyl residue . . . . .	10
1.6.	AdoMetDC reaction mechanism . . . . .	12
1.7.	Topology of human AdoMetDC . . . . .	13
1.8.	Dimer interface of human AdoMetDC . . . . .	14
1.9.	Known AdoMetDC inhibitors . . . . .	15
1.10.	Putrescine-binding and active-sites in the human enzyme . . . . .	16
2.1.	A brief overview of homology modelling . . . . .	22
2.2.	Predicted ORFs from <i>P. berghei</i> and <i>P. yoelii</i> . . . . .	25
2.3.	AdoMetDC fragments from <i>P. chabaudi</i> and <i>P. knowlesi</i> . . . . .	25
2.4.	Homology of human secondary structural elements . . . . .	26
2.5.	Final modelling alignment . . . . .	27
2.6.	Predicted secondary structures of AdoMetDC insert 2 . . . . .	28
2.7.	Consensus secondary structure predictions for AdoMetDC insert 3 . . . . .	29
2.8.	Ramachandran plots of the initial and final models . . . . .	30
2.9.	Topology of the final model . . . . .	31
2.10.	C- $\alpha$ trace superimposition of model on human template . . . . .	31
2.11.	Important model-active site residues . . . . .	32
2.12.	Active site substitutions . . . . .	33
2.13.	Ligand interactions for the model and human active sites with MeAdoMet . . . . .	34
2.14.	Active site shapes of AdoMetDC . . . . .	35
2.15.	Cavity near the sulphonium group . . . . .	36
2.16.	Charge network associated with insert 1 . . . . .	37
3.1.	Putrescine charge networks . . . . .	47
3.2.	Orientation of putrescine . . . . .	52
3.3.	<i>Xba</i> I and <i>Hind</i> III restriction of AdoMetDC/ODC . . . . .	53
3.4.	<i>Xba</i> I and <i>Hind</i> III restriction of mutants . . . . .	53
3.5.	Effect of mutations on AdoMetDC activity. . . . .	54

4.1. Orientations of the top 6 potential NCI inhibitors . . . . .	62
4.2. Overview of effect of identified inhibitors and solvent controls on AdoMetDC activity . . . . .	64
A.1. Multiple alignment . . . . .	87
A.2. Conserved motifs . . . . .	90

## List of Tables

1.1. Effects of key mutations in human AdoMetDC . . . . .	11
2.1. Secondary structure prediction algorithms applied to <i>Plasmodium</i> AdoMetDC/ODC sequences . . . . .	24
2.2. Backbone deviations . . . . .	30
2.3. Model and human active site composition . . . . .	33
3.1. AdoMetDC/ODC mutant primers . . . . .	49
3.2. Residues associated with putrescine stimulation . . . . .	52
3.3. Average relative activities . . . . .	54
4.1. Potential AdoMetDC inhibitors . . . . .	63
5.1. Summary of main differences between the malarial (model) and host enzymes. . . . .	68
B.1. LUDI BIOSYM virtual screening hits . . . . .	91
B.2. ACD virtual screening hits . . . . .	92
B.3. NCI virtual screening hits . . . . .	93

## Typographical conventions

- Computer related abbreviations and terms are given in PROGRAM CODE (usually uppercase) type in order to distinguish them from wet-bench and biological terms.
- Residues are referred to using the standard three letter code followed directly by the residue number of the organism in question. The organism follows directly in italics: *hum*: *Homo sapiens*, *pot*: *Solanum tuberosum* (potato). For example Ser68*hum* would refer to serine 68 of the human enzyme. When no species is given in the residue name or in the text, *P. falciparum* is assumed.
- Amino acid substitutions and mutations are indicated using the standard three letter code for the original residue, followed directly by it's position, which is in turn followed by the replacement amino acid, e.g.: Ser68Ala would indicate the replacement of serine 68 with alanine.



## List of Abbreviations

AdoMetDC	S-adenosylmethionine decarboxylase
CGP4884A	4-amidinoindan-1-one-2'-amidinohydrazone
CPM	Counts per minute
DFMO	$\alpha$ -Difluoromethyl ornithine
DDT	Dichlorodiphenyltrichloroethane
DHFR-TS	Dihydrofolate reductase-thymidylate synthase
DHPS	Dihydropteroate synthase
DMF	Dimethyl formamide
DMSO	Dimethyl sulphoxide
DNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EC	Enzyme Commission
EDTA	Ethylene diamine tetra-acetic acid
Glc6PD-6PGL	Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase
kb	Kilo base
LB	Luria-Bertani
MAOEA	5'-deoxy-5'-[N-methyl-N-[(2-aminooxy)ethyl]amino]adenosine
MeAdoMet	Methyl ester of S-adenosylmethionine
MGBG	Methylglyoxal bis(guanylhydrazone)
MHZPA	5'-deoxy-5'-[N-methyl-N-(3-hydrazinopropyl)amino]adenosine
MW	Relative molecular mass
NMR	Nuclear Magnetic Resonance
ODC	Ornithine decarboxylase
ORF	Open Reading Frame
PCR	Polymerase chain reaction
Pfu	<i>Pyrococcus furiosus</i>
PLP	Pyridoxal-5-phosphate
PMSF	Phenylmethanesulphonyl fluoride
PVL	Pyruvoyl
RMSD	Root Mean Square Deviation

SDS	Sodium dodecylsulphate
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris-HCl	Trishydroxy (methyl-amino) methane / Hydrochloric acid
Wt	Wild-type

## List of Computer Related Terms

ACD	Available Chemicals Directory
BLAST	Basic Local Alignment Sequence Tool
CFF	Consistent force field
CHARMM	Chemistry at HARvard Molecular Mechanics
CLUSTALX	Cluster Alignment (for X windows)
EMBL	European Molecular Biology Laboratory
EMBOSS	European Molecular Biology Open Source Software
FASTA	Fast Alignment
GONNET	Amino acid substitution matrix
GRID	Program from the DOCK suite for generating scoring grids
LIGPLOT	Free program for automatically plotting protein-ligand interactions
MEME	“Multiple Em (Expectation maximisation) for Motif Elicitfication”
MODELLER	Homology modelling based on satisfaction of spatial restraints
NCI	National Cancer Institute (USA)
PAM	Point accepted mutation amino acid substitution matrix
PASS	Prediction of Activity Spectra for Substances
PDB	Protein Data Bank
PERL	Practical extraction and report language
PHRAP	“phragment assembly program” for assembling overlapping DNA segments into contiguous stretches
PLASMODB	<i>Plasmodium</i> genome database
PROCHECK	A useful protein structure validation program
PYMOL	Molecular graphics viewer implemented in PYTHON
SWISS-MODEL	Server for homology modelling
SWISS-PROT	High quality annotated database of protein sequences

## Summary

Malaria affects nearly 500 million people every year. The constant evolution of resistance to existing therapies calls for the identification of new drugs and strategies to fight this disease. One way to facilitate this is the characterisation of novel parasite metabolic pathways and their exploitation. The bifunctional S-adenosylmethionine decarboxylase/Ornithine decarboxylase (AdoMetDC/ODC) enzyme, represents one such target. Within this enzyme reside the two main regulatory activities for the biosynthesis of polyamines. Furthermore, the bifunctional arrangement does not occur in the human host, and is presently unique to *Plasmodium*. This uniqueness therefore represents a potential target for the identification of new *Plasmodium*-specific drugs.

The exploitation of parasitic drug targets can be aided immensely by knowledge of its atomic 3D structure. However, malarial proteins are often reluctant to yield to traditional experimental methods for gathering this information. In this study, a computational approach was followed to gain further insight into the structure of the AdoMetDC domain of the bifunctional enzyme. The AdoMetDC domain was modelled on X-ray crystal structure templates of the human and plant equivalents.

The model revealed a number of differences compared to the human structure. Amino acid substitutions and active site shape differences suggest this enzyme is worthwhile exploiting for the discovery of new drugs. The model also revealed possible reasons for the lack of putrescine stimulation, as seen in humans, and suggested a possible replacement mechanism in the form of internal residues assuming the putrescine's function. The presence of such a replacement mechanism was partially verified experimentally by site-directed mutagenesis and recombinant expression of mutant enzymes.

The model was also used to conduct *in silico* screens against databases of small molecules for the identification of potential inhibitors. Some of these compounds were subsequently subjected to preliminary screening with recombinantly expressed enzyme. No promising inhibitors were found, however, the results provided insights for further inhibitor identification.

## Opsomming

Malaria affekteer nagenoeg 500 miljoen mense per jaar. Die konstante evolusie van weerstand-biedendheid teenoor bestaande terapeutiese middels noodsaak die identifisering en karakterisering van unieke parasietpadweë. Die bifunksionele S-adenosylmetionien dekarboksilase/Ornitien dekarboksilase (AdoMetC/ODC) proteïen verteenwoordig een so 'n teiken. Die bifunksionele ensiem verteenwoordig die twee hoof regulatoriese aktiwiteite vir die biosintese van poliamiene. Verder kom die bifunksionele rangskikking nie voor in die menslike gasheer nie, en is tans uniek tot *Plasmodium*. Hierdie unieke kenmerk verteenwoordig a potensiële teiken vir die identifisering van nuwe *Plasmodium*-spesifieke geneesmiddels.

Die ontwikkeling van parasiet geneesmiddeltekens word aansienlik bevorder deur die kennis van drie-dimensionele atoomstrukture. Malaria proteïene is dikwels moeilike tekens vir tradisionele eksperimentele metodes om hierdie inligting te bekom. In hierdie studie is 'n rekenaargesteurde benadering gevolg om verdere insig in die struktuur van AdoMetDC van die bifunksionele proteïen te bekom. Die AdoMetDC domein is gemodelleer op grond van die kristalstruktuur template van die menslike en plant ekwivalente.

Die model het 'n aantal verskille opgewys in vergelyking met die menslike struktuur. Aminosuur substitusies en vormverskille in die aktiewe setel dui aan dat die ensiem waarskynlik geskik is vir ontwikkeling van nuwe geneesmiddels. Die model het ook 'n moontlike verklaring gebied vir die afwesigheid van putresien stimulasie, soos wat by mense aangetref word, en het gedui op 'n moontlike vervangende meganisme in die vorm van interne residue wat die funksie van putresien oorneem. Die teenwoordigheid van so 'n vervangingsmeganisme is gedeeltelik eksperimenteel bevestig deur middel van setel-gerigte mutagenese en rekombinante uitdrukking van mutante ensieme.

Die model is ook gebruik om *in silico* sifting teen kleinmolekuul databasisse uit te voer, met die oog op die indentifikasie van nuwe potensiële inhibitore. Sommige van die middels is daarna gebruik vir voorlopige toetsing teenoor die rekombinante ensiem. Geen belowende inhibitore is gevind nie, alhoewel, die resultate verskaf insig vir verdere inhibitor identifikasie.



## Chapter 1

# Introduction

### 1.1. The need for new anti-malarials

Malaria is a disease characterised particularly by severe febrility. The disease acquired its name from the Latin “mal’aria” or “bad air”, after an association gained in Roman times for being prevalent in marshy areas. Writings of high fevers left by Hippocrates and others indicate that malaria has been known for at least 2500 years. The discovery of swollen spleens in Egyptian mummies suggest that the figure is closer to 5000 years. With a rapidly growing human population the disease quickly established itself over much of the Old World. European colonisers in turn carried the disease to the New World, and by the beginning of the 1900s the disease had established itself as far north as Siberia (Sherman, 1998).

Malaria is caused by mosquito-borne parasitic protozoa of the genus *Plasmodium*. The four known species capable of infecting humans are *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*. Of these, *P. falciparum* is the most infective and is responsible for the largest number of deaths annually (Miller *et al.*, 2002). The mosquito hosts are the females of the genus *Anopheles*. The most effective *P. falciparum* transmitters are *A. gambiae* and *A. funestus* which are prevalent through the African tropics (Breman, 2001). The parasite exhibits a complex life cycle that is shared between the human and mosquito hosts (Fig. 1.1). Infection begins with a mosquito bite, injecting sporozoites into subcutaneous tissue or blood which in turn infect the liver. There the parasites mature into merozoites, which later infect the erythrocytes. Within the red blood cells the parasite passes through various further developmental stages which result in the asexual reproduction of the merozoites. Once the new merozoites are produced the red blood cell ruptures releasing the merozoites for further invasion. The life cycle is completed with the maturation of merozoites into gametocytes. These are taken up by the mosquito host where sexual reproduction occurs. The asexual stage is the main pathological stage of the parasite’s life cycle (Miller *et al.*, 2002). Red blood cell rupture results in the release of parasite and erythrocyte material. It is this release, and the host reaction to these products that largely give rise to the disease. Furthermore, the loss of red blood cells can lead to anaemia (White, 1998).

The first successful treatment for malaria was quinine, derived from the bark of the *Cinchona* tree from South America. Discovered by Spanish colonists in the 17th century, quinine was brought back to Europe and was rapidly established as the prime therapeutic for malaria. Its value had a number

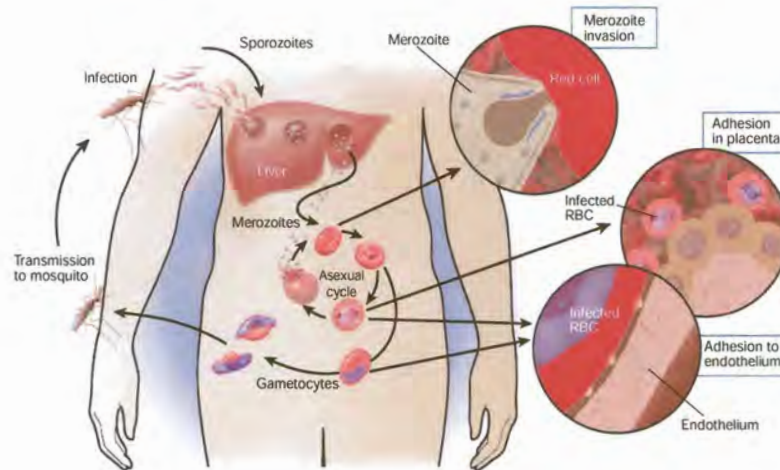


Figure 1.1: The life cycle of *P. falciparum*. Disease results from the asexual/red blood cell stage of the life cycle.

of important historical spinoffs. Attempts to synthesise quinine played a key role in the founding of the modern chemical industry (Garfield, 2000). Quinine's discovery also helped lay the foundation for rational drug discovery, with the search for other anti-malarials ultimately leading to the synthesis of chloroquine and the sulpha-drugs (Meshnick, 1998). For much of the early twentieth century malaria was largely eliminated in the industrialised world. This was mainly due to the availability of effective drugs such as chloroquine, and aided a great deal by the large scale deployment of DDT (Dichlorodiphenyl-trichloroethane) insecticide in the temperate regions (Ridley, 2002). However, for new drugs, resistant strains have consistently arisen within the first 30 years of initial deployment (Meshnick, 1998). This is true for what was the frontline treatment for many years - chloroquine, as well as the second most important treatment of sulphadoxine/pyrimethamine combined. Resistance has so far been seen to emerge in the lower transmission areas such as South America and South East Asia. A couple of reasons have been suggested for this. Firstly, the lower immunity of individuals living in these areas may give the parasite a greater chance of evolving resistance when drugs are given. Secondly, these individuals, being less immune, are considered more likely to seek medical intervention (Meshnick, 1998). Nonetheless, resistance continues to emerge, ultimately spreading to sub-Saharan Africa where the greatest burden exists (Ridley, 2002)(Fig. 1.2). This is in turn compounded by the emergence of insecticide resistant *Anopheles* vectors (Ridley, 2002).

A number of anti-malarial drugs do exist. However, existing drug resistance or the possible emergence thereof limits their usefulness. A brief overview of current anti-malarial drugs is given below:

- Quinine and analogues: The exact mechanism of action of quinine and its analogues is still much in debate. The consensus is that anti-malarial action seems to be due to interference of these compounds with the correct functioning of the food vacuole and haem metabolism (Krogstad and Dibyendu, 1998).
  - \* Quinine: While still effective against many resistant strains, resistance is becoming more frequent. However, required regimens often result in poor patient compliance. Associated toxicity can also





Figure 1.2: Current global status of malaria resistance. Red areas indicate malaria endemic regions. Figure from Ridley (2002).

lead to hearing loss, tinnitus and nausea. Quinine is often also administered in combination with antibiotics for it to be fully effective (Meshnick, 1998).

- \* Chloroquine: Cheap to produce and very effective against non-resistant malaria, this drug proved to be the mainstay of malarial therapy for much of the first half of the twentieth century. However, the now near worldwide resistance of the parasite to this drug has rendered it useless for the present (Ridley, 2002).
- \* Mefloquine: Useful as a weekly prophylactic as well as for treatment. However, resistance to this drug emerged quickly, and mefloquine is known to have neurologic side effects (Milhous and Dennis, 1998).
- \* Halofantrine: Found to be effective where mefloquine fails, although resistance is not unknown. However, the high doses that are required for effective treatment also lead to cardiotoxicity (Milhous and Dennis, 1998).
- \* Amodiaquine: Found to be effective against most chloroquine resistant strains. However, prophylaxis has been found to cause agrunalarcytosis (Milhous and Dennis, 1998).
- \* Lumefantrine: A relatively new drug given in combination with artemether, an analogue of the artemisinin family. Although related to mefloquine, no neurotoxicity has yet been reported (Milhous and Dennis, 1998).
- Antifolates: The antifolates target pathways that the parasite uses for nucleotide and amino acid biosynthesis.
  - \* Pyrimethamine/Sulphadoxine: Pyrimethamine inhibits dihydrofolate reductase (DHFR) which reduces dihydrofolate to tetrahydrofolate. This enzyme is a useful target in that it is common both to *de novo* tetrahydrofolate synthesis, as well as exogenous folate salvage pathways. Sulphadoxine is a sulphonamide that inhibits the action of dihydropteroate synthase (DHPS), which is required

earlier in the tetrahydrofolate biosynthesis pathway. Resistance to this dual combination has unfortunately also spread rapidly (Fig. 1.2) (Cowman, 1998).

- \* Cycloguanil: This compound is the active metabolite of proguanil, and also inhibits DHFR. Proguanil has been found to be very effective in combination with atovaquone, an inhibitor of mitochondrial electron transport. In particular, development of resistance to the latter is prevented by the inclusion of proguanil. However, the cost of this combination has made it prohibitive (Krogstad and Dibyendu, 1998; Vaidya, 1998).
- Artemisinin and analogues: Artemisinin was discovered in China in 1967 to be the active ingredient of the *Artemisia annua* wormwood. In order to improve its pharmacological profile a number of derivatives have been developed. This drug has been recently identified to target a  $\text{Ca}^{2+}$  ATPase from *P. falciparum* although a number of other mechanisms have been suggested (Eckstein-Ludwig *et al.*, 2003). Field resistance to this drug has yet to emerge, although resistance has been induced in the laboratory. These drugs are usually administered with other anti-malarials due to their short half-lives (Meshnick, 1998).

Malaria is therefore still a serious global concern, owing to the consistent development of drug resistance. Approximately 300-500 million acute cases of malaria occur annually. Between 1 and 3 million deaths result, with 90% of the burden existing in sub-Saharan Africa (Bremner, 2001). Specifically for South Africa this has required the reintroduction of DDT spraying for vector control (Hargreaves *et al.*, 2000). This problem is compounded by a lack of new anti-malarials. Between 1975 and 1996 only 3 out of 1223 new drugs entering the market were anti-malarials (Greenwood and Mutabingwa, 2002). The recent arrival of the artemisinin based drugs should alleviate the burden for a while. However, the ability of the parasite to evolve resistance to new drugs means efforts must be made to constantly expand our arsenal (Ridley, 2002).

Recent efforts, however, raise hopes that the malaria threat can be dealt with. The completion of the genome sequence for *P. falciparum* has already aided in the identification of potential drug targets and associated drugs (Gardner *et al.*, 2002). Among these was the identification of genes associated with type II fatty acid synthesis. This target is particularly attractive as type II fatty acid synthesis is not present in humans. Thilactomycin which targets a number of fatty acid II enzymes in *E. coli* has been shown to inhibit *P. falciparum* growth. Another compound, Triclosan was demonstrated to be effective *in vitro* in rodent models after identification of the fatty acid II enzyme enoyl-acyl-carrier protein reductase from preliminary sequence data (Ridley, 2002; Hoffman *et al.*, 2002). Isoprenoid biosynthesis was also identified as a potential drug target from preliminary genome data. As a result the known isoprenoid inhibitor fosfidomycin successfully demonstrated anti-malarial activity *in vitro* against *P. falciparum*, and *in vivo* against the rodent malaria *P. vinkei* (Hoffman *et al.*, 2002). Other potential targets include parasite specific proteases, glucose transport, glycolysis and various targets within the apicoplast, an organelle that is believed to be of bacterial origin (Gleeson, 2000). The presence of this organelle has



meant that a number of existing prokaryotic translation and transcription inhibitors such as doxycycline, clindamycin and tetracycline can be used effectively against malaria (Ridley, 2002). Some of these drugs are only effective on their own as prophylactics, or have to be administered in combination with other drugs for curative purposes. This study will focus on polyamine metabolism, which is also considered to be a potential drug target.

## 1.2. Polyamines

### 1.2.1. Functions of polyamines

The polyamines are a class of polycationic molecules characterised by multiple amine groups. The most important of these are putrescine, spermidine and spermine (Fig. 1.3). Polyamines have so far been found to be present in all organisms and are always required for normal physiological functioning (Tabor and Tabor, 1985). Their requirement is particularly emphasised in cells undergoing rapid proliferation (Marton and Pegg, 1995).

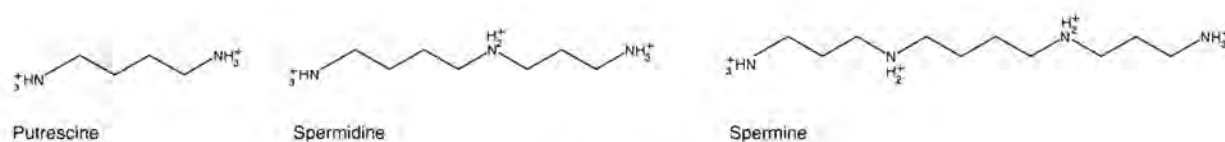


Figure 1.3: The main polyamines.

Although only a few specific requirements have been identified for these molecules, their main function is still thought to be the physical stabilisation of DNA and RNA. This is expected to be mediated by electrostatic interactions that can take place between the cationic nitrogens and the polycationic nucleotide backbones (Igarashi *et al.*, 1982; Tabor and Tabor, 1984a). This is further validated by the observation that in certain biological systems where polyamine biosynthesis is interrupted, an accumulation of putrescine is observed, that in turn may assume the function of the larger polyamines. This only seems to be effective at much higher concentrations of putrescine to spermidine and spermine (Marton and Pegg, 1995). Only a few specific biosynthetic functions have been identified for polyamines. Spermidine is required for the posttranslational modification of eukaryotic translation initiation factor eIF-5A to the hypusinated form. Hypusine is a non-translated amino acid formed by the post-translational modification of lysine, using spermidine as a substrate. It has been suggested that this function of spermidine is the main reason for the effectiveness of polyamine inhibitors in cancer cells (Byers *et al.*, 1992, 1994). Spermidine is also required by the Kinetoplastids, for example parasites of the *Trypanosoma* genus, responsible for such diseases as African Sleeping Sickness and Chagas disease. In this case spermidine is needed for the synthesis of trypanothione, a spermidine and glutathione conjugate that is involved in maintaining the redox balance. Trypanothione essentially consists of two glutathione molecules linked by a spermidine bridge. Its function in maintaining the redox balance is essentially similar to glutathione, and works



by the cyclical formation/breaking of an internal disulphide bond (Müller *et al.*, 2003). Spermine itself appears to have little function, except possibly to stimulate certain mitochondrial uptake processes. It may also serve as a polyamine store because it can be converted back to spermidine and putrescine (Marton and Pegg, 1995).

### 1.2.2. Polyamine metabolism

The polyamine pathways described here resembles the well regulated mammalian system. Metabolic interconversions and feedback mechanisms allow the cell to respond to loss or gain of polyamines. Specifically, polyamine products inhibit ODC and AdoMetDC activity, whereas putrescine stimulates AdoMetDC activity in some organisms, including mammals (Marton and Pegg, 1995). The generic pathway is outlined in Fig. 1.4. Putrescine is produced either by ornithine decarboxylase (ODC, EC 4.1.1.17) or from arginine by the consecutive actions of arginase and agmatine ureohydrolase. The putrescine product serves as a scaffold for donation of aminopropyl groups from decarboxylated S-adenosylmethionine (dcAdoMet). dcAdoMet is produced by S-adenosylmethionine decarboxylase (AdoMetDC, EC 4.1.1.50) (Tabor and Tabor, 1984a,b).

ODC and AdoMetDC have been identified as the main rate-limiting enzymes of polyamine biosynthesis, and hence the main targets of inhibitory studies (Marton and Pegg, 1995). The addition of the aminopropyl groups is carried out by spermidine synthase and spermine synthase, respectively. These two enzymes are the next most important in regulating the polyamine pool. The aminopropyl donation is essentially irreversible. However, spermine and spermidine can be converted back to their precursors by the sequential action of *N*<sup>1</sup>-acetyltransferase and polyamine oxidase. The latter enzymes can act on the parent polyamines, albeit more slowly. The activities of both ODC and AdoMetDC are inhibited by high polyamine content. Inhibition of polyamine metabolism is usually not sufficient to remove all polyamines from the metabolic pool. Most organisms that have been studied are able to obtain exogenous polyamines via uptake by membranous transport proteins (Marton and Pegg, 1995). Furthermore, mammalian ODC and AdoMetDC have very short half-lives, (10-20 min ODC, 20 min - 2hr AdoMetDC), amongst the shortest of any proteins. This allows for rapid turnover of the key enzymes in polyamine biosynthesis which can quickly negate any inhibition (Tabor and Tabor, 1984a). As described above, ODC and AdoMetDC activity are down-regulated by the higher polyamine end-products that result from their activity. Hence low polyamine levels stimulate ODC and AdoMetDC activity in order to compensate for the depleted polyamine pool (Marton and Pegg, 1995). Inhibition of ODC leads to increased AdoMetDC activity due to the low amounts of the dcAdoMet cosubstrates, putrescine and spermidine. Conversely, inhibition of AdoMetDC leads to increased ODC activity, as evidenced by the large observed increase in putrescine (Marton and Pegg, 1995). As a result complete inhibition of ODC and AdoMetDC is difficult to attain. ODC activity and expression are in turn regulated by the protein antizyme. Apart from inhibiting ODC, antizyme binds to ODC in order to facilitate recognition by the proteasome for degradation. Antizyme acts catalytically in this regard, and can be recycled for degradation of more than

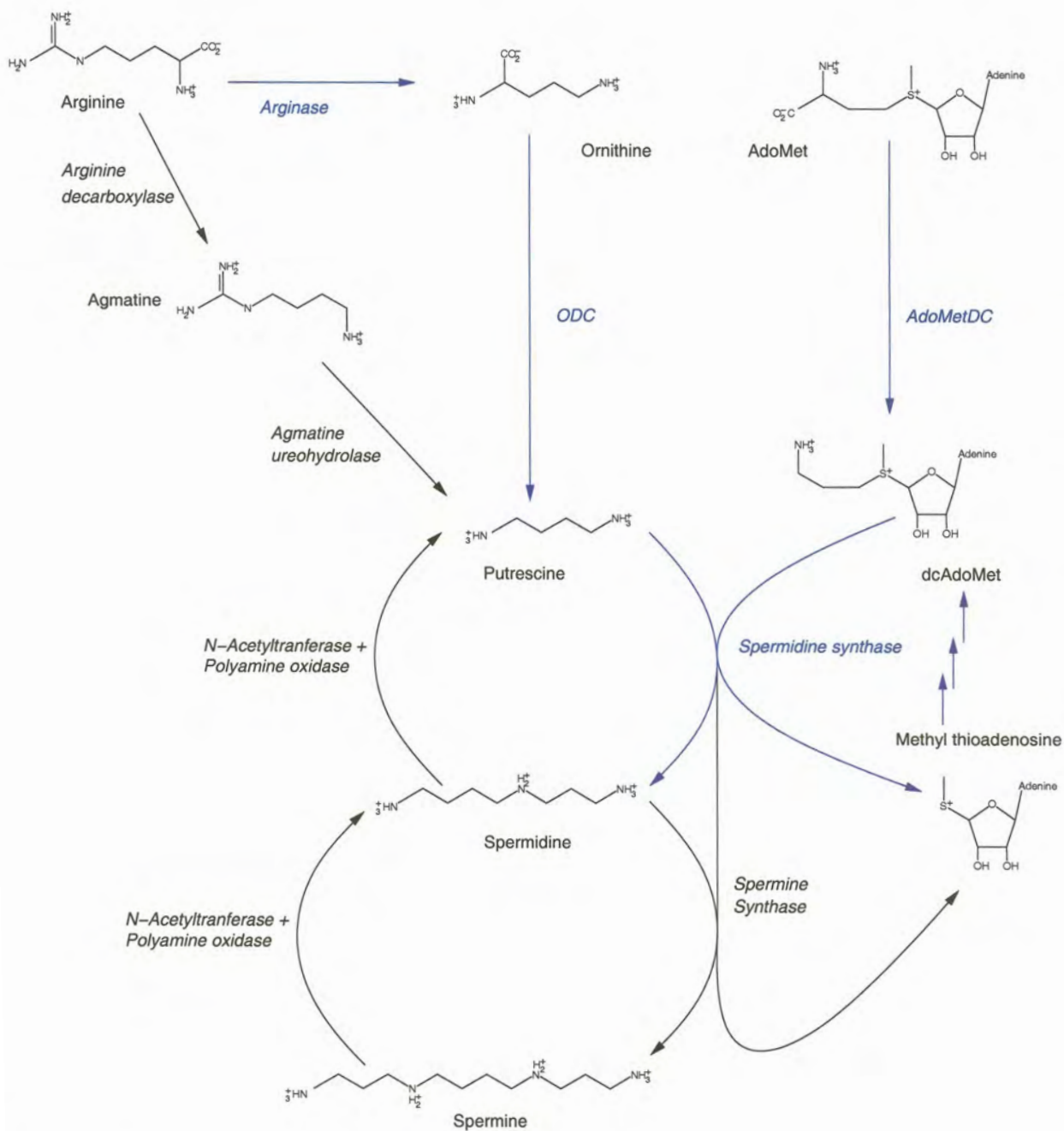


Figure 1.4: The generic polyamine pathway based on the mammalian system. Pathways and enzymes coloured blue indicate those that have been found in *P. falciparum*. ODC: Ornithine decarboxylase, AdoMetDC: S-adenosylmethionine decarboxylase.



one ODC molecule. This is in contrast to most proteins which are targeted for proteasome degradation by the covalent addition of ubiquitin. Polyamines in turn regulate antizyme, by inducing a translational frame shift during antizyme expression (Coffino, 2000).

### 1.2.3. Polyamines in malaria

The corresponding pathways in *P. falciparum* have not yet been fully elucidated. Nonetheless, key differences have already emerged compared to mammalian polyamine metabolism. In most organisms ODC and AdoMetDC activities reside in separately expressed proteins. Eukaryotic ODC is an obligate dimer with both subunits contributing residues to each active site (Seely *et al.*, 1982; Almrud *et al.*, 2000). Human AdoMetDC is also observed to form a dimer, however, each active site exists wholly within one subunit (Ekstrom *et al.*, 1999). In *P. falciparum* however, ODC and AdoMetDC domains reside in a single bifunctional protein complex of 330 kDa (Müller *et al.*, 2000). In order for the malarial ODC activity to exist, two proteins ( $\pm 160$  kDa each) must in turn associate to form the functional complex. *P. falciparum* also possesses spermidine synthase and arginase. The remaining notable differences are the apparent absence of spermine synthase and the  $N^1$ -acetyltransferase and polyamine oxidases required to convert spermidine and spermine to their precursors. Transport has been observed for putrescine and spermidine (Müller *et al.*, 2001). Spermine transport is assumed due to an apparent lack of spermine synthase and an increase in erythrocyte spermine levels upon infection with *P. falciparum* (Assaraf *et al.*, 1987). There is some similarity in ODC and AdoMetDC regulation from *P. falciparum* in that spermidine exhibits weak inhibition of these two enzymes. The effect is greater for putrescine, with a more marked effect on malarial ODC. Malarial AdoMetDC is not stimulated by putrescine (Wrenger *et al.*, 2001), which is in contrast with the human enzyme (Pegg, 1984). An antizyme-like regulation system for the parasite has yet to be reported, and otherwise appears to be absent.

### 1.2.4. Polyamines as a drug target

The dependence of rapidly proliferating cells on polyamines, has meant that polyamine biosynthesis has been investigated for some time as an anticancer target (Marton and Pegg, 1995). A number of inhibitors for the main enzymes (ODC, AdoMetDC, spermine- and spermidine synthase) have been identified as a result (Byers *et al.*, 1992, 1994; Wang, 1995). No successful anticancer drugs have come from these efforts, however. Polyamine inhibition is usually unable to completely deplete the polyamine pool and therefore tends to induce cytostasis rather than cytotoxicity (Marton and Pegg, 1995). For the reasons outlined above (Section 1.2.2) it is generally difficult to kill target cells by polyamine inhibition. This would suggest that in order to target polyamine metabolism for therapeutic purposes, multiple enzymes and/or the transporters would have to be targeted. This has been successfully demonstrated using polyamine analogues that inhibit cellular polyamine uptake, in conjunction with the irreversible ODC inhibitor  $\alpha$ -difluoromethyl ornithine (DFMO) in breast carcinoma cells (Graminski *et al.*, 2002). It is

also intriguing to consider the possibility that a single molecule could potentially inhibit multiple enzymes, because the polyamine structural motif presents itself a number of times in polyamine metabolism.

The state of affairs is somewhat different for *P. falciparum* however. The short ODC and AdoMetDC half-life is not observed for the bifunctional malarial enzyme, suggesting that the parasite would be more susceptible to targeting of these enzymes (Wrenger *et al.*, 2001). The extended half-life and the bifunctional nature of AdoMetDC/ODC are key features that distinguish malarial polyamine metabolism from the mammalian host. This identifies malarial polyamine metabolism as a potential drug target, and by exploiting these differences it may be possible to identify malaria-specific drugs. The short turnover of the mammalian enzyme in turn suggests that there is a good chance of discovering novel anti-malarials with favourable pharmacological profiles, since host polyamine metabolism is less likely to be undesirably perturbed.

Investigations into the potential anti-malarial activity of DFMO show that it has little effect on the erythrocytic stages of the parasite using the *P. berghei* rodent model, and that *in vitro* activity against *P. falciparum* is also cytostatic rather than cytotoxic (Bitonti *et al.*, 1987). This may possibly be due to poor uptake of the compound and/or the ability for the parasite to utilise exogenous polyamines (Müller *et al.*, 2001). Whatever the reason, it is likely that any anti-polyamine strategy that is followed for malaria will have to deal with the transport problem. Inhibition of ODC by the ornithine analogue DFMO has already been successful for the treatment of West African Sleeping Sickness caused by *T. brucei gambiense*. For DFMO to be used effectively it must be given intravenously in large doses. It is also unfortunately ineffective against *T. brucei rhodesiense* (East African Sleeping Sickness, Wang, 1995). Targeting of the Trypanosomal AdoMetDC has also been successfully demonstrated *in vitro* and *in vivo* in mice. The AdoMetDC inhibitor CGP 40215A inhibited at a  $K_i$  of 4.5 nM and was found to successfully cure *Trypanosoma* infected mice when used in combination with DFMO (Brun *et al.*, 1996; Bacchi *et al.*, 1996). These results suggest that pursuing polyamine metabolism for intervention in parasitic diseases may be worthwhile. The properties and inhibition of AdoMetDC will be discussed in further detail in the following section.

### 1.3. Properties of S-adenosylmethionine decarboxylase (AdoMetDC)

#### 1.3.1. AdoMetDC requires pyruvoyl

Eukaryotic AdoMetDC is a pyruvoyl-requiring enzyme, usually  $\pm$  330 amino acids in length. Unlike most decarboxylases which require pyridoxal-5'-phosphate (PLP) for activity, AdoMetDC falls among a small class of enzymes that use a covalently bound pyruvoyl instead. Other enzymes that make use of pyruvoyl include aspartate decarboxylase, histidine decarboxylase, proline reductase and phosphatidyl serine decarboxylase (Marton and Pegg, 1995). The pyruvoyl group is derived during an internal proteolytic cleavage from a serine residue (Ser68 - *H. sapiens*) (Stanley *et al.*, 1989). In the eukaryotic



AdoMetDC family this residue resides within a conserved -ESS- motif (converted residue underlined). Studies carried out on the human enzyme revealed that the cleavage is autocatalytic and non-hydrolytic (Recsei and Snell, 1984; Tabor and Tabor, 1984a). Processing yields two subunits, the larger C-terminal  $\alpha$ -chain (38.4 kDa) and the smaller N-terminal  $\beta$ -chain (7.7 kDa). The processing reaction is a serinolysis, whereby the serine residue that is converted attacks the carbonyl carbon of the preceding peptide bond to form an oxyxazolidine intermediate (Fig. 1.5). This undergoes an N $\rightarrow$ O acyl shift to form an ester intermediate,  $\beta$ -elimination follows to yield dehydroalanine and glutamate. The dehydroalanine tautomerises to the imine form, and is then hydrolysed to ammonia and pyruvoyl, thus generating the pyruvoyl residue on the N-terminus of the  $\alpha$ -subunit. Both a proton donor and acceptor are required for this process. Mutational evidence indicates that Ser229 and His243 assume these functions, respectively (Xiong *et al.*, 1999). Mutating Ser68 to alanine produces an inactive, unprocessed enzyme (Xiong *et al.*, 1997). Further mutation studies indicate Ser229 may be the proton acceptor required for the first step of processing, possibly increasing the nucleophilicity of the attacking -OH of Ser68 (Xiong and Pegg, 1999). Ser229Ala does not process, and Ser229Cys processes slowly, whereas Ser229Thr processes normally. Treatment of the His243Ala mutant with the base hydroxylamine accelerated cleavage, which otherwise occurred very slowly. This indicates that His243 serves as the base needed to abstract hydrogen from the Ser68  $\alpha$ -carbon during the  $\beta$ -elimination step (Tolbert *et al.*, 2001). A summary of important mutations and their effects is given in Table 1.1.

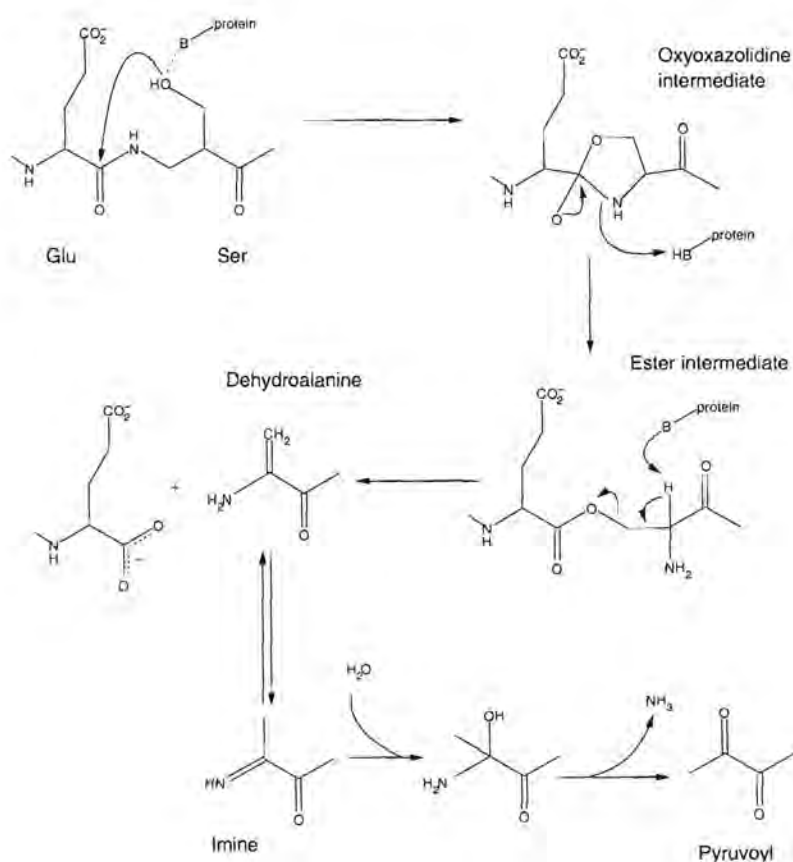


Figure 1.5: Formation of the AdoMetDC pyruvoyl residue (Bennett *et al.*, 2002).



Table 1.1: Effects of key mutations in human AdoMetDC.

Residue	Mutant	Processing	Activity	Function	Ref.
Ser68	Ala	Eliminated	Eliminated	Converted to pyruvate	(Stanley <i>et al.</i> , 1989) (Tolbert <i>et al.</i> , 2003b)
	Cys	Slowed	Converted to thio-carboxylate		(Xiong <i>et al.</i> , 1997)
	Thr	Slowed	Converted to $\alpha$ -ketobutyrate		(Xiong <i>et al.</i> , 1997)
Cys82	Ala	Slower, less stimulation by putrescine	Eliminated	Protonation of carbonyl oxygen during first step of processing. Protonation during decarboxylation	(Stanley and Pegg, 1991) (Tolbert <i>et al.</i> , 2003b)
	Ser	Slower	Not specified		(Tolbert <i>et al.</i> , 2003b)
Glu8	Gln	Normal	Eliminated		(Stanley and Pegg, 1991)
Glu11	Gln	Not stimulated by putrescine	Eliminated	Required for putrescine stimulation of processing	(Stanley and Pegg, 1991)
	Asp	Inhibited by putrescine	Eliminated		(Xiong <i>et al.</i> , 1999)
	Lys	Eliminated	Eliminated		(Stanley <i>et al.</i> , 1994)
Lys80	Ala	Not stimulated by putrescine	Substantially reduced		
Glu178	Gln	Not stimulated by putrescine	Minimal Not stimulated by putrescine	Required for putrescine stimulation of processing and activity	(Stanley <i>et al.</i> , 1994)
Glu256	Gln	Not stimulated by putrescine	Substantially reduced Not stimulated by putrescine	Required for putrescine stimulation of processing and activity	(Stanley <i>et al.</i> , 1994)
Tyr112	Ala	Eliminated	Eliminated		(Stanley <i>et al.</i> , 1994)
Glu133	Gln	Putrescine absolutely required for processing	Almost eliminated		(Stanley <i>et al.</i> , 1994)
Asp174	Asn	Not stimulated by putrescine	Not stimulated by putrescine	Required for putrescine stimulation of processing and activity	(Xiong <i>et al.</i> , 1997)
His243	Ala	Incomplete processing Enzyme trapped in ester intermediate	Eliminated	Proton abstraction during $\beta$ -elimination of ester intermediate	(Xiong <i>et al.</i> , 1999) (Ekstrom <i>et al.</i> , 2001)
	Glu	Slower	Eliminated		(Xiong <i>et al.</i> , 1999)
Ser229	Ala	Eliminated	Eliminated	Possible proton donor during processing	(Xiong and Pegg, 1999) (Tolbert <i>et al.</i> , 2003b)
	Cys	Very slow	Eliminated	May be required for product release during catalysis	(Xiong and Pegg, 1999) (Tolbert <i>et al.</i> , 2003b)
	Thr	Normal	Eliminated	May increase nucleophilicity of carbonyl carbon during processing	(Xiong and Pegg, 1999) (Tolbert <i>et al.</i> , 2003b)
Phe7	Ala	Normal	Substantially reduced	Required for correct binding of substrates and inhibitors	(Tolbert <i>et al.</i> , 2001)
Phe223	Ala	Normal	Substantially reduced	Required for correct binding of substrates and inhibitors	(Tolbert <i>et al.</i> , 2001)

### 1.3.2. Enzymatic mechanism

Initial biochemical and mutational evidence involving human AdoMetDC was consistent with the original hypothesis that the pyruvoyl residue functions as an electron sink during the decarboxylation reaction. This facilitates the weakening of the C-C<sub>α</sub> bond, allowing the carboxyl on the α-carbon of S-adenosylmethionine to become a leaving group. The consensus reaction mechanism is outlined in Fig. 1.6. A Schiff base is formed between the pyruvoyl moiety and the amino group of S-adenosylmethionine, much as with enzymes employing PLP, where a similar Schiff base is formed between the cofactor and substrate molecule. The α-carbon is then reprotoneated and the resulting Schiff base hydrolysed to release the product (Allen and Klinman, 1981; Ekstrom *et al.*, 1999). Reprotonation is most likely carried out by Cys82, seeing as mutating Cys82 to alanine results in an inactive enzyme (Xiong *et al.*, 1999). Glu11 has also been identified as important for enzyme activity. Mutation of Glu11 to Asp or Gln reduces activity and stimulation by putrescine (Stanley and Pegg, 1991; Xiong and Pegg, 1999).

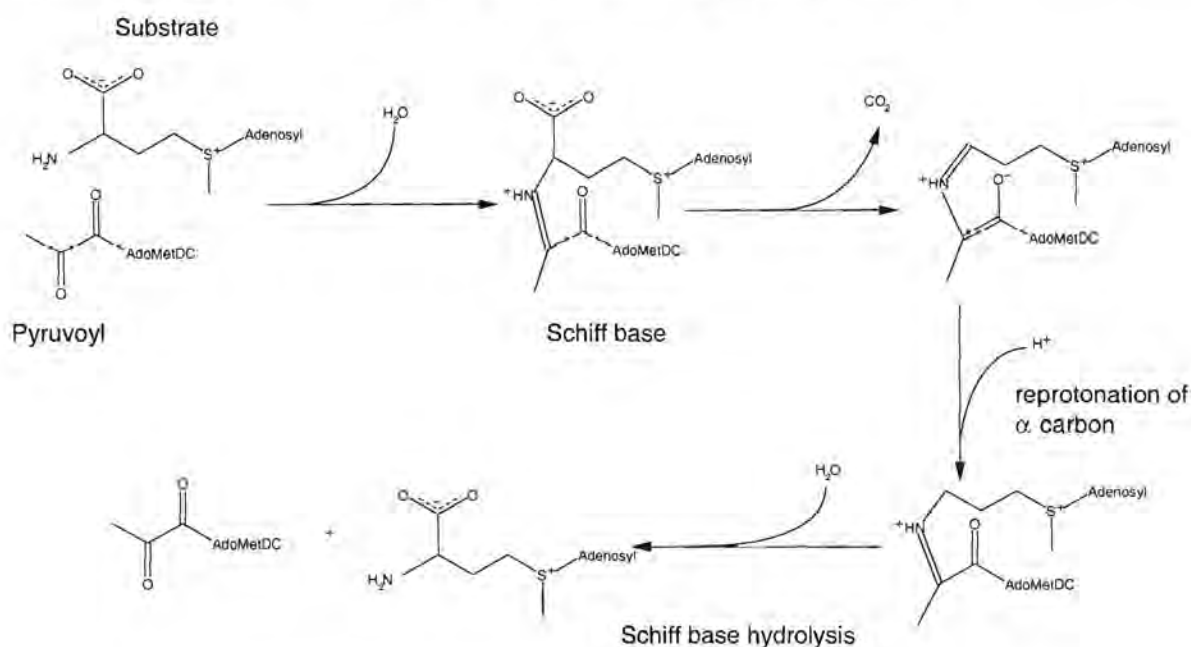


Figure 1.6: AdoMetDC reaction mechanism (Tolbert *et al.*, 2001).

#### 1.3.2.1. Effects of putrescine on AdoMetDC

In certain organisms putrescine, and in some cases other polyamine analogues, stimulate AdoMetDC. This effect can be either on activity or processing and varies from species to species. However, less is known about the effect of putrescine on processing than on activity from the various organisms that have been studied. Both activity and processing are stimulated in the human enzyme (Stanley and Pegg, 1991). In plants (Xiong *et al.*, 1997) and *Plasmodium* (Wrenger *et al.*, 2001) no effect is observed on enzyme activity or processing. In the fungus *Neurospora crassa* putrescine is not seen to exert an effect on processing, although it is absolutely required for enzyme activity (Hoyt *et al.*, 2000). In baker's yeast (*Saccharomyces cerevisiae*) AdoMetDC activity is also stimulated by putrescine, however,



the effect on processing is unknown (Pösö *et al.*, 1975b). Increased activity has also been established for the protozoic parasite *Trypanosoma*, however, no effect is observed on processing. Closer analysis reveals that *Trypanosoma* have only limited elements of putrescine stimulation that have been identified in humans (Persson *et al.*, 1998; Kinch *et al.*, 1999). In the free living protozoan *Tetrahymena pyriformis* no putrescine stimulation of AdoMetDC activity is observed (Pösö *et al.*, 1975a). This is also observed in the slime mould *Physarum polycephalum*, where putrescine has no effect on AdoMetDC activity (Mitchell and Rusch, 1973). No case where putrescine stimulates processing but not activity has yet been reported.

### 1.3.3. Structure of AdoMetDC

#### 1.3.3.1. The AdoMetDC fold

The structures of the human and potato enzymes have been determined by X-ray crystallography to resolutions of 2.25 Å (Ekstrom *et al.*, 1999) and 2.3 Å (Bennett *et al.*, 2002), respectively. From these studies the AdoMetDC fold was seen to comprise two layers of  $\beta$ -sheets stacked between two layers of  $\alpha$ -helices. The active site lies between the two  $\beta$ -sheet layers of the  $\alpha\beta\beta\alpha$ -sandwich near the surface of the enzyme (Fig. 1.7). Each  $\alpha\beta$ -half (from here on referred to as an  $\alpha\beta$ -slice) of the protein has a similar topology and is thought to be the result of an ancient gene duplication (Ekstrom *et al.*, 1999). The human enzyme functions as a dimer (Stanley *et al.*, 1989) and crystallises as such with an edge-on interaction between  $\beta$ -strands B7 and B15 from each monomer (Fig. 1.8). The potato enzyme does not dimerise (Bennett *et al.*, 2002).

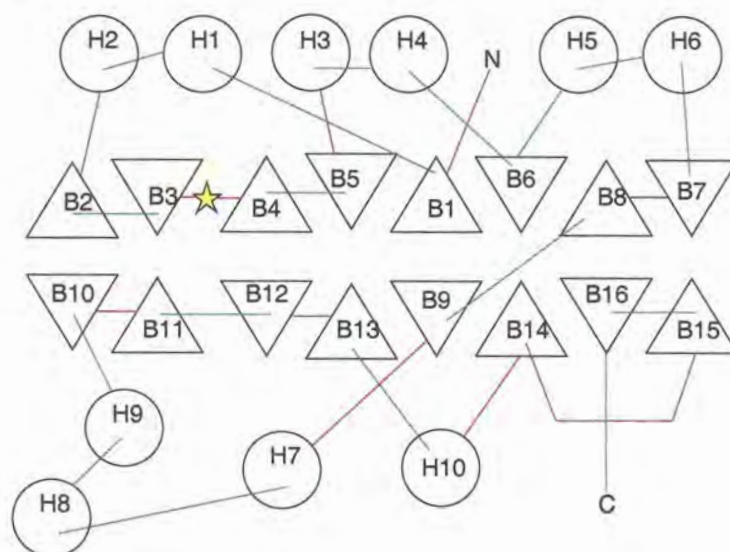


Figure 1.7: Topology of human AdoMetDC. Circles represent  $\alpha$ -helices, triangles  $\beta$ -sheets. The active site and site of proteolytic cleavage is designated by the yellow star. N: N-terminus, C: C-terminus. Anterior loops are coloured green, and posterior loops magenta.

The AdoMetDC structures so far determined bear little resemblance to those of other pyruvoyl requiring enzymes, namely histidine decarboxylase from *Lactobacillus* (Gallagher *et al.*, 1993), aspartate decarboxylase from *E. coli* (Albert *et al.*, 1998) and arginine decarboxylase from *Methanococcus jan-*

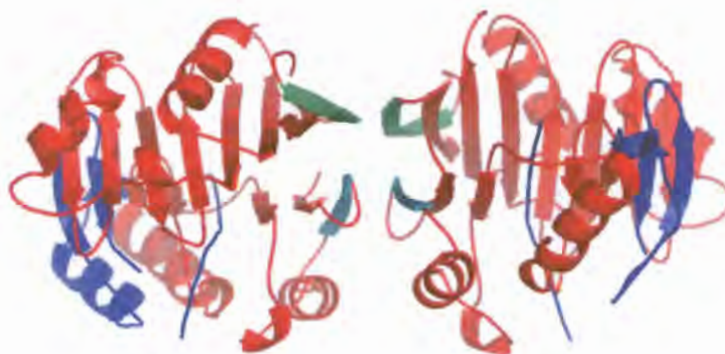


Figure 1.8: Dimer interface of human AdoMetDC. The  $\alpha$ -subunits are coloured red, and the  $\beta$ -subunits blue. Green  $\beta$ -strand 15. Cyan  $\beta$ -strand 7. Generated in PYMOL. PDB ID: 1JEN.

*naschii* (Tolbert *et al.*, 2003a). Arginine decarboxylase and histidine decarboxylase do bear some structural similarities. In both structures the  $\alpha$ - and  $\beta$ -subunits associate to form trimers. Two of these trimers in turn associate to form a hexamer in histidine decarboxylase. Similar to AdoMetDC, in both histidine decarboxylase and arginine decarboxylase the monomer has an  $\alpha\beta\beta\alpha$  arrangement of helix- and  $\beta$ -strand-layers. However, the topology of the arginine- and histidine decarboxylase monomers is different to that of AdoMetDC. The cleavage that generates pyruvoyl in these two enzymes and aspartate decarboxylase, occurs in a loop connecting two  $\beta$ -strands from separate  $\beta$ -sheets. In AdoMetDC this occurs between two strands in the same  $\beta$ -sheet. Furthermore, the active sites of arginine- and histidine decarboxylase are comprised of amino acids donated from two subunits. In contrast, the active sites of AdoMetDC dimer are located distinctly in each monomer (Tolbert *et al.*, 2003a). In contrast to most  $\beta$ -rich proteins, only one linker region is observed to link each  $\alpha\beta$ -half of AdoMetDC. Two such loops are observed in the  $\alpha\beta\beta\alpha$ -sandwich arrangements of arginine- and histidine decarboxylase.

### 1.3.3.2. AdoMetDC structure, processing and enzyme activity

A series of human AdoMetDC crystal structures encompassing the wild-type and two mutants has allowed the sequence of events that occurs from unprocessed pro-enzyme to functional AdoMetDC to be explored (Ekstrom *et al.*, 1999, 2001; Tolbert *et al.*, 2003a). The Ser68Ala mutant remains completely unprocessed with the peptide bond between residues 67 and 68 still intact. The Ser68 residue is thus confirmed as the residue that is converted to pyruvoyl. The Ser68Ala structure was mutated *in silico* to the original wild-type form (Tolbert *et al.*, 2003b). Molecular modelling of this mutated structure identified a number of residues involved in processing that had been previously implicated in mutagenesis studies (Table 1.1). From this it was suggested that the hydroxyl of Ser68 may attack the previous carbonyl carbon without a hydrogen donor to increase nucleophilicity (Fig. 1.5). The hydrogen donor role has already been proposed for Ser229, however, it is now considered possible that bond strain may facilitate this reaction. Ser229 and/or Cys82 were found to be close enough to form hydrogen bonds with the carbonyl carbon of Glu67, thus possibly increasing its electrophilicity towards Ser68. Modelling of



an oxyoxazolidine intermediate has indicated that Ser229 is also close enough to donate a proton to the resulting oxyoxazolidine anion of the ring intermediate, and thus stabilise this structure. The structure of the His243Ala mutant reveals an unprocessed protein trapped in the ester intermediate before the N→O acyl shift (Fig. 1.6). His243 is close enough to donate a proton to the N atom of the oxyoxazolidine ring, and may thus assist the N→O acyl shift to form the ester during processing.

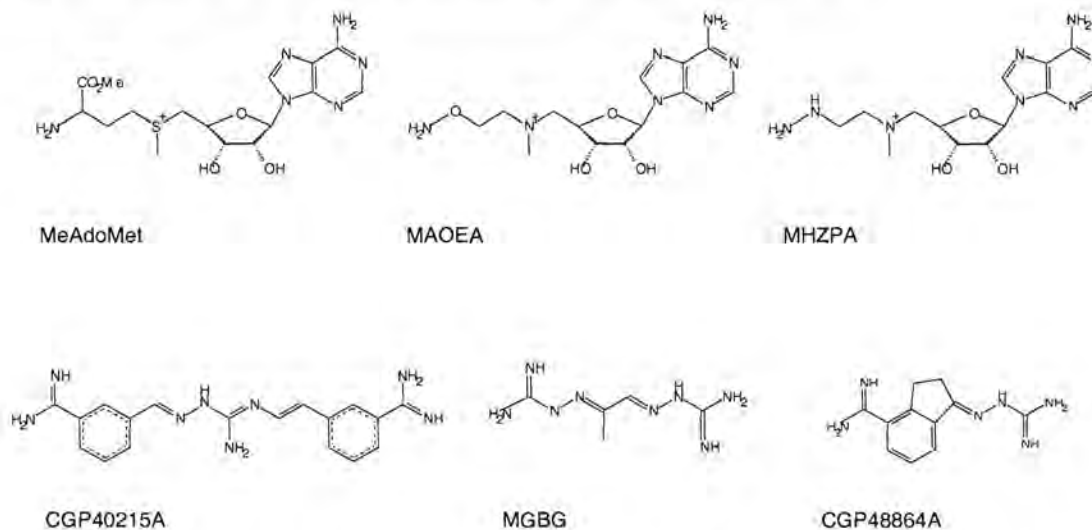


Figure 1.9: Known AdoMetDC inhibitors. MeAdoMet: S-adenosyl methyl ester, MAOEA: 5'-deoxy-5'-[N-methyl-N-[(2-aminooxy)ethyl]amino]adenosine, MHZPA: 5'-deoxy-5'-[N-methyl-N-(3-hydrazinopropyl)amino]adenosine, MGBG: Methylglyoxal bis(guanylhydrazone), CGP48864A: 4-amidinoindan-1-one-2'-amidinohydrazone.

Crystal structures of human AdoMetDC with various known inhibitors have also been obtained (Fig.1.9, Tolbert *et al.*, 2001). From this a number of important interactions could be inferred about the binding of the natural substrate to AdoMetDC. AdoMetDC inhibitors fall into two classes: substrate-analogue inhibitors and non-substrate-analogue competitive inhibitors. The substrate analogue inhibitors include MeAdoMet (methyl ester of the natural substrate), MHZPA (5'-deoxy-5'-[N-methyl-N-(3-hydrazino-propyl)amino]adenosine) and MAOEA (5'-deoxy-5'-[N-methyl-N-[(2-aminooxy)ethyl]amino] adenosine), which bind irreversibly to the enzyme by formation of a Schiff base with the pyruvoyl residue. The two hydroxyl groups of the ribose moiety from these analogues each form a hydrogen bond to one of the oxygens of the side-chain carboxyl of Glu247. The adenine moiety adopts the syn conformation, and is hydrogen bonded to Glu67. Furthermore, the adenine ring is stacked between the phenyl moieties of Phe7 and Phe223. This stacking is also observed for the planar regions of the competitive inhibitors CGP48864A (4-amidinoindan-1-one-2'-amidinohydrazone) and MGBG (Methylglyoxal bis[guanylhydrazone]). Mutation of either of these residues to alanine results in decreased inhibition, indicating that aromatic stacking interactions are required for normal binding of the substrate and inhibitors. Each of the inhibitors tested carries a positive charge, in agreement with the positive sulphonium ion that exists in the substrate. However, no negatively charged residues have been observed close to this group, even though inhibitors which lack a positive charge are found to be less effective (Pankaskie and Abdel-Monem, 1980; Pegg and Jacobs, 1983). The methyl ester modification on the  $\alpha$ -carboxyl in the

MeAdoMet crystal structure prevents this moiety from becoming a leaving group (Tolbert *et al.*, 2001). Out of the three crystal structures with substrate analogues, this structure conserves the largest number of atoms present in the native reaction. It was therefore possible to examine the role of residues required for catalysis. From this Cys82 is seen to be close enough to protonate the decarboxylated Schiff base allowing for pyruvate regeneration. His243 may also serve this function, although it is predicted to do so at a lower frequency. Thr245 is hydrogen bonded to a water molecule that forms part of a hydrogen bond network involving Gly9, Glu11 and His243 and another water molecule. The water molecules may mediate Cys82 protonation, since replacement of Thr245 with alanine reduces activity (Xiong *et al.*, 1999).

### 1.3.3.3. AdoMetDC structure and putrescine stimulation

The ODC product putrescine is seen to stimulate the processing of human AdoMetDC by approximately one order of magnitude (Stanley *et al.*, 1994). Crystal structures of human AdoMetDC reveal that putrescine binds at an allosteric site about 20 Å away from the active site (Fig. 1.10). The mechanism of putrescine stimulation is not known. It has been suggested that putrescine binding induces a relative shift in the two  $\beta$ -sheets of the  $\alpha\beta\beta\alpha$ -sandwich. In addition, a number of residues that have been observed to affect putrescine stimulation after mutation, appear to form a network of alternating charges connecting the allosteric and active sites (Ekstrom *et al.*, 2001; Tolbert *et al.*, 2001).

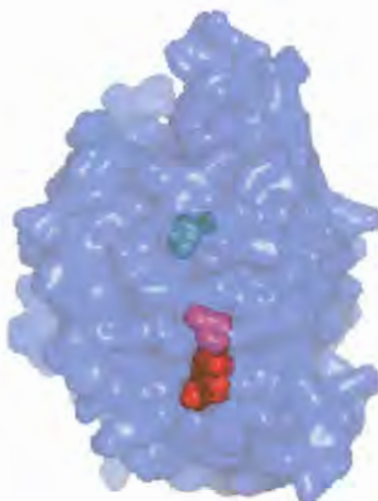


Figure 1.10: Putrescine-binding and active-sites in the human enzyme. The putrescine binding site (green) is approximately 20 Å from the active site (docked with MeAdoMet, red). Generated in PYMOL. PDB ID: 1I7B.

The positive amine groups of putrescine are seen to interact with the acidic residues Glu178, Glu256, Asp175 and Glu15, either directly or via water molecules. One end of putrescine is stabilised by Asp174 and Glu15, while the other is stabilised by Glu178, Glu256 and Glu15. The positive residue Lys80 interacts with two of these residues, Glu178 and Glu256 as well as with Glu11 which is not located near putrescine, but is found rather in the active site. The Lys80 residue has a poorly resolved electron



distribution indicating that it may be flexible. It has therefore been suggested that putrescine may exert its effect via this series of charged residues. Glu11 may be required for the  $\beta$ -elimination step of processing, possibly mediating the deprotonation of the  $\alpha$ -carbon by His243. The binding of putrescine may therefore result in a shift in the position of Glu11 in order to mediate deprotonation. Putrescine may also bring about a shift in the relative positions of the  $\beta$ -sheets of the  $\alpha\beta\beta\alpha$ -sandwich, which brings the residues required for processing into the correct position. The positive charges of the amine groups may be required to neutralise the charges on the buried residues Asp174, Glu178, Glu256 and Glu15 (Ekstrom *et al.*, 2001).

Similar suggestions have been made for the effect of putrescine on enzyme activity. Mutating Glu11 results in a catalytically inactive enzyme and is hence required for catalysis (Stanley and Pegg, 1991). Therefore a similar set of electrostatic interactions as described above may be required for the correct functioning of Glu11. The proposed  $\beta$ -sheet shift may also be required for correct alignment of residues involved in enzyme activity. It has also been suggested that such a shift may lower the  $K_m$  of AdoMetDC for its substrate. No crystal structure of the enzyme without putrescine has been obtained. However, such a shift becomes apparent when comparing AdoMetDC structures with substrate analogues compared to those with competitive inhibitors and the unliganded structure. It may therefore be that the binding of putrescine induces a similar shift which is more favourable to substrate binding (Tolbert *et al.*, 2001).

Neither the activity nor processing of AdoMetDC from plants are stimulated by putrescine (Xiong *et al.*, 1997). The crystal structure of potato AdoMetDC reveals a number of possible reasons for the lack of putrescine stimulation (Bennett *et al.*, 2002). Firstly, a number of amino acid substitutions are present which are expected to prevent the binding of putrescine. Some of these substitutions are in turn suggested to take over the role of putrescine. Most notably Arg18*pot* and Arg114*pot* (Leu and Phe in humans, respectively) occupy a region where an amine of putrescine would be expected. Furthermore, a similar network of charged residues as described for the human enzyme connects these arginines to the active site. The case is similar in *Trypanosoma* where mutating the Arg13 residue to Leu abolishes putrescine stimulation of activity, and the presence of the connecting charged network can be inferred from multiple sequence alignments (Clyne *et al.*, 2002).

#### 1.3.4. Malarial AdoMetDC

As mentioned in section 1.2.3, AdoMetDC in *P. falciparum* coexists with ODC in a bifunctional enzyme of approximately 160 kDa. Two of these proteins in turn associate to form a 330 kDa complex that results in a heterotetramer after cleavage as the functional enzyme (Müller *et al.*, 2000). The exact delineation of the domains has yet to be biochemically determined. It is known that the AdoMetDC domain occupies approximately the first 570 residues of the N-terminus and is connected by a hinge region to the ODC domain that resides approximately within the last 600 residues of the C-terminus. The catalytic activities of the two domains appear to have no regulatory effect on each other (Wrenger *et al.*, 2001). The domains themselves are much larger than their counterparts in other organisms, due

to the presence of *Plasmodium*-specific inserts (Müller *et al.*, 2000; Birkholtz *et al.*, 2003). Homology modelling of the ODC domain has revealed the presence of two such inserts. Further deletion mutagenesis studies have revealed that these inserts are required for the correct functioning of their respective domains. Furthermore, removal of inserts from one domain was also shown to decrease activity in the other domain. It was therefore suggested that inter- and intra-domain interactions resulting from the malaria-specific regions are required for normal functioning of the bifunctional enzyme. Inter-domain regulation within this bifunctional complex cannot therefore yet be ruled out (Birkholtz *et al.*, 2003, 2004). In summary a number of properties of malaria AdoMetDC conspire to make this enzyme unique: the bifunctional organisation with ODC, the *Plasmodium* specific inserts and the lack of putrescine stimulation. These differences make it a viable proposition to be exploited for novel inhibitor identification.

#### 1.4. Aims

The hypothesis of this study was that a homology model of the *P. falciparum* AdoMetDC domain could be used to guide experimental analysis. This study aimed to gain insight into the three dimensional (3D) structure of AdoMetDC from *P. falciparum* chiefly through *in silico* methods. The structural modelling portion was followed up with biochemical investigations in order to test predictions made using the model regarding residues that may be important to the enzyme's normal functioning, and the binding of novel inhibitors. It is anticipated that this knowledge would contribute to the identification of novel anti-malarials targeted specifically against polyamine metabolism. The specific objectives were as follows:

- **Chapter 2 - Structural modelling of AdoMetDC from *P. falciparum*:** The objective was to obtain a model that could be used with confidence to guide initial experiments to probe the structure and functioning of malarial AdoMetDC.
- **Chapter 3 - Model guided mutational analysis of malarial AdoMetDC:** The structural model was used to guide site directed mutagenesis of recombinantly expressed AdoMetDC/ODC. This was done in order to determine the effect on the enzyme's functioning. The results in turn give an indication as to the correctness of the model and whether it can be reliably used for further experiments.
- **Chapter 4 - Model guided inhibitor screening of malarial AdoMetDC:** The model was used to screen libraries of small molecules *in silico* in order to identify potential novel inhibitors. Some of these inhibitors were in turn selected to be tested biochemically. This once again could be used as to indicate if the model was correct. Furthermore, good inhibitors identified in this manner may prove to be potential lead compounds for novel drugs.

In the following chapter the modelling of malarial AdoMetDC is described. The properties of the model are discussed and analysed in order to gauge it's potential usefulness for further studies.



## Chapter 2

# Structural modelling of *P. falciparum* AdoMetDC

## 2.1. Introduction

### 2.1.1. The need for Bioinformatics

Genome sequencing initiatives have generated a vast amount of information that utterly precludes manual analysis. The current EMBL nucleotide database contains over 27 million sequences (<http://www.ebi.ac.uk/embl/>, Stoesser *et al.*, 2003). The SWISS-PROT database of annotated protein sequences contains over 122 000 entries (<http://www.ebi.ac.uk/swissprot/>, Boeckmann *et al.*, 2003). The challenge is to make intelligent use of this information to direct biochemical experimentation and ultimately gain holistic knowledge of how organisms function. This glut of data has occasioned the rise of Bioinformatics, which can be broadly used to group all disciplines that employ computational methods to make these incredibly large datasets manageable in order to gain biologically relevant information.

This chapter concerns the application of some of these techniques to further understand a malarial protein. To fully understand an enzyme it is necessary to have a 3D model of the protein in question. This allows the possibility of new potential inhibitors to be identified in a more rational approach instead of biochemical screening against a random library of compounds. Experiments for probing the mechanisms and functioning of enzymes can also be guided by structural knowledge. Known inhibitors can also be rationally modified and tested *in silico* in order improve properties of a potential drug, most notably substrate-enzyme binding (Böhm and Klebe, 1996; Krumrine *et al.*, 2003). The most reliable 3D protein models are the products of X-ray crystallography and NMR investigations (Flower, 2002). The Protein Data Bank (PDB) currently contains approximately 14 000 structures of proteins determined using X-ray diffraction and NMR (The PDB Team, 2003). This is far less than the number of known proteins, and shows that many more protein structures are needed. Hence the use of computational modelling to fill this gap. Furthermore, malarial proteins have proved difficult to crystallise. The malarial genome is extremely A+T-rich ( $\pm 80\%$ , Gardner *et al.* 2002), which results in drastically altered codon usage, and consequently it is difficult to express malarial proteins in crystallisable yields in the most common heterologous systems (Hyde *et al.*, 1989; Withers-Martinez *et al.*, 1999). Secondly, the frequent presence of *Plasmodium* specific inserts tends to render proteins resistant to crystallisation. These inserts tend to be of low complexity, and dominated by hydrophilic residues, especially Asn and Lys (Pizzi and Frontali,

2001). For these reasons it is often necessary to follow other methods to determine the structures of malarial proteins. In this study computational methods are used to study the structure of AdoMetDC from *P. falciparum*.

### 2.1.2. Computational protein modelling

The most reliable method for computer modelling of a molecule is to determine its quantum mechanical description. Due to the large computational resources required, the quantum mechanical description can only be determined for small molecules (a few hundred atoms) and is generally still not feasible for molecules the size of proteins. Liu *et al.* (2001) have demonstrated that quantum mechanical simulation is possible using a supercomputer and a semi-empirical description for the protein, crambin. Even though quantum mechanics provides the most accurate answers, another approach must be followed for large molecules like proteins. Most often this takes the form of classical Newtonian mechanics. In a mechanical treatment the molecule is split up into a number of geometrical components such as bonds, bond-angles and torsions, etc. The energy of each component is modelled and included in a large sum describing the molecule that can be referred to as the scoring or energy function. The collection of mathematical forms that is used to describe each geometrical component is referred to as the force field. For example bond stretching may be modelled by Hooke's law for a spring:  $U(r_{ab}) = \frac{1}{2}k_{ab}(r_{ab}-r_{ab-eq})^2$ , where  $U$  represents the energy,  $r_{ab}$  the radius between atoms  $a$  and  $b$ ,  $r_{ab-eq}$  the radius at equilibrium and  $k_{ab}$  is a force constant (Cramer, 2002). Initially such force fields were designed to give a reasonable physical description of the molecule in question. However, it is also possible to include terms that bear no relation to physical reality, the rationale being that if they produce a reliable model, the lack of resemblance to reality is not important. One such method is the MODELLER algorithm (Šali and Blundell, 1993). The MODELLER scoring functions include terms from the CHARMM forcefield, as well as terms describing the probabilities of geometrical components. These probabilities are derived from databases of known structures of proteins, and will describe for instance the distribution of each dihedral angle in residue side chains, etc.

Once a scoring function is obtained for a protein, it can be put to a number of uses. In order to understand a protein it is desirable to know its structure in the native state. This is often assumed to be the lowest energy conformation of the protein, because it is the most likely (Anfinsen, 1973). This assumption is not always valid, because proteins are flexible and are better understood as occupying an ensemble of low energy conformations. However, it is nonetheless useful to be able to predict these states (Cramer, 2002). The scoring function that is used to represent a protein typically describes a complex multidimensional potential energy surface. In order to find the low energy conformations, systematic adjustment of the molecule's coordinates is needed in order to find the minimum values of such a function (minimisation). The potential energy surface is usually too complex to be explored exhaustively. However, a number of methods such as steepest descents and conjugate gradients, can be followed to find a good approximation of a minimum. The details of these methods will not be entered



into here, suffice to say that with any model some degree of minimisation is required to ensure that undesirable (high energy) geometric conformations are removed (Jensen, 1999; Cramer, 2002).

Even though the mechanical approximation of a protein is more tractable, it would still require too much computing time to model a large protein from its linear extended conformation *ab initio*. Thus simply knowing the protein's amino acid sequence and the structures of the amino acids themselves is not sufficient. Therefore a good starting point is required before minimisation techniques can be applied. The most common method is knowledge-based or homology modelling. Using this technique a model can be built using template structures which have been previously solved for other organisms. These structures are usually of the same protein, i.e. performing the same function in a different organism. However, if the target protein is known to fall within a particular fold, e.g. the triosephosphate isomerase barrel, a template protein of different function but the same topology can be used. The most important step in this process is the alignment between the amino acid sequences of the target protein and template protein/s. This step is based on the assumption that the target and template sequences have diverged from a common ancestor protein. Chothia and Lesk (1986) demonstrated that highly diverged proteins often fall into the same fold. This was further confirmed by Sander and Schneider (1991) who found that for all known protein structures, all sequences of more than 100 amino acids, with 30% or greater sequence identity, were structurally similar. This rule has recently been revised using a larger set of known structures: structural similarity can exist below 20% sequence identity, however, it is also possible for structures with greater than 30% identity to be structurally dissimilar (Rost, 1999). The information within the alignment determines what template residue is used to model a particular target residue (Fiser *et al.*, 2000). There are a number of algorithms which can be employed for homology modelling, e.g. the SWISS-MODEL (Schwede *et al.*, 2003) server and MODELLER (Šali and Blundell, 1993). The MODELLER algorithm also makes use of probability distributions for various geometric properties such as torsion angles and bond lengths, generated from large libraries of existing structures in order to constrain the model (satisfaction of spatial restraints, Fig 2.1). The MODELLER program was used in this study since previous attempts to use the SWISS-MODEL server had proved unsuccessful due to low sequence identity between the target and template (Dr Birkholtz, Personal communication).

Homology modelling has been successfully used in number of cases, including malaria proteins. As mentioned in section 1.2.3 the ODC domain of AdoMetDC/ODC from *P. falciparum* has been modelled (Birkholtz *et al.*, 2003) as has triosephosphate isomerase (Joubert *et al.*, 2001). Notably, homology modelling of the bifunctional DHFR-TS enzyme of *P. falciparum* has been successfully used to explain drug resistance (Rastelli *et al.*, 2000) and design new inhibitors that are effective in the nanomolar range (McKie *et al.*, 1998; Yuthavong *et al.*, 2000).

The main objective of the study outlined in this chapter was to construct a homology model of the AdoMetDC domain of the bifunctional AdoMetDC/ODC from *P. falciparum*. The fitness of this model is then discussed, as well as other properties of the enzyme that were determined in the process of model construction.

i17343057  
b16338

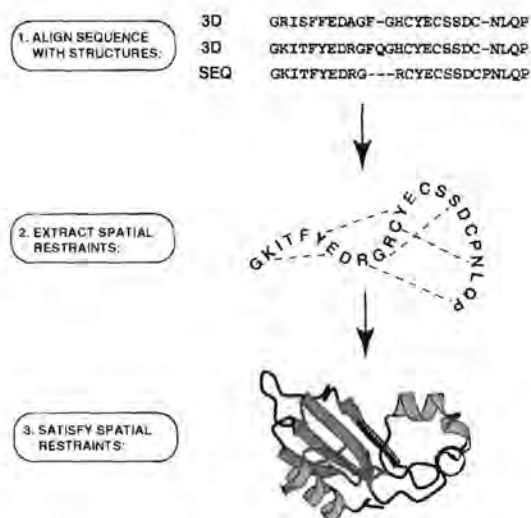


Figure 2.1: A brief overview of homology modelling (MODELLER manual). Firstly, the target sequence is aligned against the template sequences. This alignment is then used to devise the set of spatial restraints that the model of the target sequence must satisfy. The energy function representing the model is then modified until those restraints are satisfied.

## 2.2. Methods

### 2.2.1. Identification of other *Plasmodium* sequences

Because the bifunctional enzyme was unique to *P. falciparum* it was of interest to determine whether other species of the *Plasmodium* genus also possessed this sequence. The full sequence of the bifunctional protein from *P. falciparum* was therefore compared to the PLASMO DB (The *Plasmodium* Genome Database Collaborative, 2001) database of *Plasmodium* sequences. This was carried out using the BLAST algorithm (Altschul *et al.*, 1990) as provided on the PLASMO DB website (<http://www.plasmodb.org>). When necessary, low-complexity filtering and repeat masking were disabled in order to try and detect the corresponding regions in other species. The searches were conducted using the database of 1/02/2002. Any fragments that appeared to belong to the same gene were assembled into contiguous stretches with PHRAP (Rieder *et al.*, 1998). Open reading frames were identified with GETORF and PLOTORF from the EMBOSS suite (<http://www.emboss.org>).

### 2.2.2. Multiple alignment

In constructing the multiple alignment for modelling purposes, other *Plasmodium* sequences identified as explained above were included. This was done in order to aid identification of conserved core regions of the enzyme as opposed to *Plasmodium* specific inserts. A large number of sequences from other eukaryotes was included in order to remove the bias introduced by the *Plasmodium* sequences. Sequences from the following organisms were used: *Bos taurus*, *Homo sapiens*, *Mesocricetus auratus*, *Mus musculus*, *Rattus norvegicus*, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Onchocerca volvulus*, *Leishmania donovani*, *Trypanosoma brucei brucei*, *Trypanosoma cruzi*, *Arabidopsis thaliana*, *Brassica juncea*, *Catharanthus roseus*, *Datura stramonium*, *Dianthus caryophyllus*, *Helianthus annuus*, *Hordeum chilense*,



*Zea mays*, *Nicotiana sylvestris*, *Oryza sativa*, *Pisum sativum*, *Pharbitis nil*, *Solanum tuberosum*, *Spinacia oleracea*, *Nicotiana tabacum* and *Saccharomyces cerevisiae* (Accession numbers are given in App. A).

Alignment was carried out with CLUSTALX 1.81 (Thompson *et al.*, 1997), using the GONNET set of scoring matrices. Gap-opening and gap-extension penalties of 15 and 0.31 were used, respectively. The delay divergent sequences property was set to 20%, and negative matrices used. Furthermore, in order to identify conserved motifs MEME 3.0.3 (Bailey and Elkan, 1994) was used on the same set of sequences. The alignment was manually adjusted in order to incorporate the high scoring identified motifs, care also being taken not to introduce any disruptions in  $\alpha$ -helices and  $\beta$ -strands.

### 2.2.3. Secondary structure prediction

In order to gain further insight into the potential structure of *P. falciparum* AdoMetDC all full length bifunctional AdoMetDC/ODC sequences were subjected to various secondary structure prediction algorithms (23 in total). The GARNIER algorithm was included with the EMBOSS suite. The other algorithms were supplied on various web servers as listed in Table 2.1. Three overlapping segments were then generated corresponding approximately to the N-terminal, middle and C-terminal regions of the proteins. Where necessary scripts written in PERL were used to convert results to FASTA files for comparison in alignment programs. The same predictions were run on DHFR-TS from *P. falciparum*, *P. vivax*, *P. chabaudi* and *P. berghei* in order to determine which methods performed better at identifying known regions of secondary structure using the *P. falciparum* crystal structure as reference (Yuvaniyama *et al.*, 2003).

### 2.2.4. Homology modelling

The templates used for homology modelling were that of the Human AdoMetDC crystal structure (1.9 - 2.7 Å, PDB entry: 1I7B) irreversibly complexed with the methyl ester of the substrate (MeAdoMet), and the crystal structure of unbound potato AdoMetDC (2.3 Å, PDB entry: 1MHM). Two templates were used in order to provide a larger set of identical template-target residue pairs. The complexed human crystal structure was used in order to generate a model more suitable for inhibitor identification, since upon inspection of the complexed versus uncomplexed crystal structures, significant shifts were observed for the C-terminus of the  $\alpha$ -chain, and the loop connecting helix 2 with  $\beta$ -strand 2. The potato crystal structure is uncomplexed, but does bear greater sequence similarity to the model in certain regions than the human sequence. Heavy atom (N,C,O) homology models were built with MODELLER 6v2 (Šali and Blundell, 1993; Marti-Renom *et al.*, 2000; Fiser *et al.*, 2000), using a high refinement (REFINE\_4) to generate 100 models. The pyruvoyl residue was modelled by manually editing the MODELLER library files. The required CHARMM residue topology for pyruvoyl was generated within INSIGHTII (Accelrys®). Ramachandran and other protein plots were generated for each of the models using PROCHECK (Morris *et al.*, 1992). From this set one model was chosen that had a minimum number of residues in disallowed regions of

Table 2.1: Secondary structure prediction algorithms applied to *Plasmodium* AdoMetDC/ODC sequences.

Program	Source
GARNIER	EMBOSS
JNETPRED JNETHMM JNETALIGN JNETPSSM JNETFREQ JPRED	<a href="http://www.compbio.dundee.ac.uk">http://www.compbio.dundee.ac.uk</a>
PHD DoublePrediction HNHC SOPM SIMP96 PREDATOR DSC Sec. Cons GOR4	<a href="http://npsa-pbil.ibcp.fr">http://npsa-pbil.ibcp.fr</a>
SAM-T99	<a href="http://www.cse.ucsc.edu">http://www.cse.ucsc.edu</a>
HMMSTER	<a href="http://www.bioinfo.rpi.edu">http://www.bioinfo.rpi.edu</a>
PROF PROFSUB PHD PHDSUB PSIPRED	<a href="http://cubic.bioc.columbia.edu">http://cubic.bioc.columbia.edu</a>

the Ramachandran plot. After superimposition on the human template, the irreversibly bound methyl ester of the substrate was transferred to this model. Hydrogens were added to the heavy atom model within INSIGHTII using a pH setting of 7.2. Certain bonds of the non-protein groups had to be manually adjusted. The adenine group of MeAdoMet was treated as a delocalised system and the O-C connections of the pyruvoyl moiety were modified to double bonds. A hydrogen, which was automatically added to the sulphonium sulphur of the bound inhibitor was removed. The formal charge of the sulphonium atom was then set to +1, and other potentials were assigned within INSIGHTII. Minimisation was performed in INSIGHTII with the CFF91 forcefield. A cutoff of 9.5 Å was used for van der Waals and electrostatic interactions. A distance dependant dielectric constant of 4 was used in order to simulate a protein environment (INSIGHTII documentation). Energy minimisation was carried out in two phases. The first phase comprised steepest descent minimisation until a maximum derivative of  $1000 \text{ kcal.mol}^{-1}.\text{Å}^{-1}$  was reached. This was followed in the second phase with conjugate gradients until a maximum derivative of  $1 \text{ kcal.mol}^{-1}.\text{Å}^{-1}$  was reached. Results of modelling were analysed using LIGPLOT 4.1.1 (Wallace *et al.*, 1995), CERIU2 (Accelrys ®) and PYMOL. The DSSP algorithm (Kabsch and Sander, 1983) was used to assign secondary structure as implemented in the *dsspcombi* program version April 1 2000.



## 2.3. Results

### 2.3.1. Identification of other *Plasmodium* sequences

For the rodent malaria species *P. berghei* and *P. yoelii*, sequences with high scores corresponding to the full bifunctional protein were identified. From *P. berghei* the berg-296a09.q1c sequence was identified from PLASMO DB with a smallest sum-probability ( $P(N)$ ) of  $1 \times 10^{-219}$ , and from *P. yoelii*, chrPyl\_cpy1465 was identified with a  $P(N)$  value of  $4.9 \times 10^{-224}$ . Smaller values indicate a more significant match, a value of zero indicates complete identity between the query sequence and the matched sequence. Open reading frame (ORF) identification indicated a large single exon gene of  $\pm 4$ kb (Fig. 2.2) for each of these genomic sequences. From *P. yoelii* a sequence of 4212 bp (1404 aa) was extracted, and from *P. berghei* a 4209 bp (1403 aa) sequence was found. Both of these genomic segments, as well as that containing the bifunctional gene for *P. falciparum*, display a similar distribution of ORFs (Fig. 2.2).

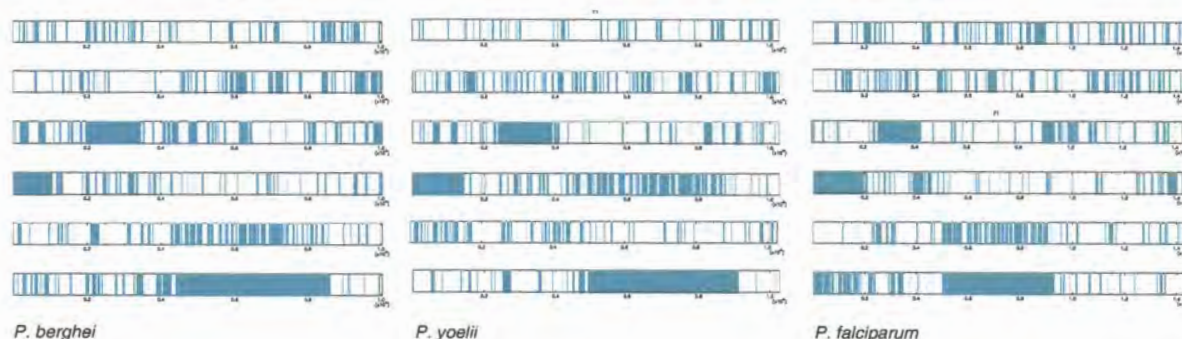


Figure 2.2: ORFs (turquoise) from the genomic regions identified from *P. berghei* and *P. yoelii*. Each row corresponds to another reading frame. All six reading frames are shown. Default parameters were used which resulted in all regions enclosed by start and stop codons. Each genomic region corresponds to  $\pm 10$ kb. The large open reading frame for each species corresponds to the bifunctional AdoMetDC/ODC sequence in each case ( $\pm 4$ kb).

For *P. chabaudi* and *P. knowlesi* a number of high scoring short fragments (30-40 residues) were found, and subsequently assembled into contiguous sequences. Three fragments were assembled for *P. chabaudi*, and two fragments for *P. knowlesi*. Both full sequences of *P. yoelii* and *P. berghei*, like *P. falciparum*, have N-termini similar to AdoMetDC sequences from other eukaryotes, and C-termini similar to other ODC sequences. Of the fragments from *P. chabaudi*, two bear resemblance to the ODC region, while the other aligns well with the AdoMetDC domain. *P. knowlesi* in turn yielded two fragments: one similar to the AdoMetDC region and the other similar to the ODC domain (Fig. 2.3).



Figure 2.3: AdoMetDC fragments from *P. chabaudi* and *P. knowlesi*. The approximate location of fragments from *P. chabaudi* and *P. knowlesi* relative to the *P. falciparum* full bifunctional sequence is shown. *Plasmodium* specific inserts are also indicated.

### 2.3.2. Alignment and motif identification

The AdoMetDC domains of *Plasmodium sp.* are highly divergent from that of other eukaryotic enzymes. From the multiple alignment (Fig. A.1, App. A) three *Plasmodium* specific inserts were obvious. The program MEME which is able to find shared patterns in unaligned sequences was used to independently identify conserved motifs. The results of motif identification confirmed the presence of three inserts (Fig. A.2, App. A). The first insert is 7 residues long and occurs in all *Plasmodium* species for which the sequence is known. There is only moderate conservation between sequences ( $\pm 43\%$  identity between *P. falciparum* and the other sequences). The second insert is longer ( $\pm 27$  residues) and displays a higher degree of similarity ( $\pm 50\%$  between *P. falciparum* and the other sequences), with a similar distribution of hydrophobic and polar residues between the *Plasmodium* species (Fig. A.1, App. A). The third and longest insert is much more divergent. In *P. falciparum* the insert is about 150 residues long, whereas in the rodent-infecting parasites it is about 100 residues longer ( $\pm 250$  residues total, Fig. A.1, App. A).

After removing the two larger inserts (insert 2 and insert 3), more of the conserved core structure became evident. Three alignments generated with GONNET, PAM150 and PAM250 scoring matrices were compared. This semi-quantitative analysis of the homology of the various secondary structure elements ( $\alpha$ -helices, 3-10-helices and  $\beta$ -strands) of the human enzyme structure reveals that the more conserved elements cluster around the active site. The more divergent elements were in turn generally located furthest from the active site. The inclusion of other *Plasmodium* sequences aided the identification of corresponding motifs between the template sequences and target sequences. Elements which had proved difficult to identify when only using the *P. falciparum* sequence were now easier to find (Fig. 2.4).

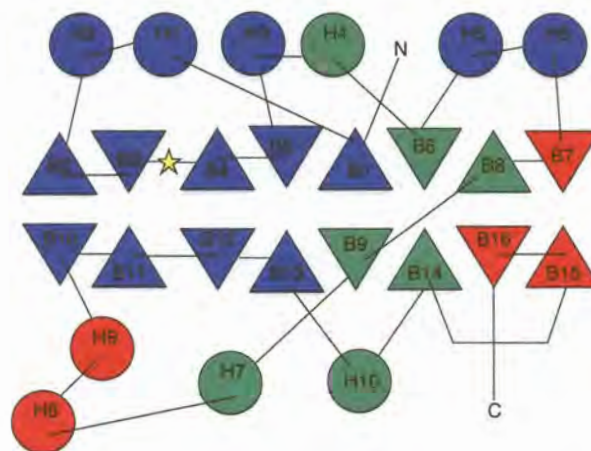


Figure 2.4: Homology of human secondary structural elements. The human topology diagram is used to indicate the degree of homology of human secondary structural elements and their predicted cognates in the malarial enzyme. Three alignments generated from GONNET, PAM150 and PAM250 scoring matrices were compared. Blue indicates agreement between all three alignments. Red indicates little agreement between any of the three alignments. Elements in green had proved difficult initially, but later proved easier to identify with the inclusion of other *Plasmodium* sequences.

A number of manual adjustments were made to the alignment. Specifically a *Plasmodium* specific insert originally from Glu55 to Arg61 was moved downstream by two residues in order to prevent an insertion within  $\beta$ -strand 2. The resulting insert (Lys57-Glu63) is defined as insert 1. Another malaria



specific 27-residue insert (Lys115-Tyr142) was moved upstream by 5 residues to relieve a disruption within  $\beta$ -strand 6. This insert is defined as insert 2 (Val110-Lys137). Due to the difficulties of reliable *de novo* modelling of loops much longer than  $\pm 12$  residues (Fiser *et al.*, 2000), the majority of insert 2 was excised, leaving overhangs of 2 residues long. The region corresponding to  $\beta$ -helix 6 and the 4 residues preceding it was realigned according to the results of MEME motif identification by moving it downstream 5 residues. This region was identified as the 5th out of 30 motifs using standard parameters. This consequently resulted in the region corresponding to  $\beta$ -strand 7 being shifted downstream by 1 residue. The short gap after Pro142 in the human sequence was shifted upstream by one residue, and corresponds to a short insert only found in the plant AdoMetDCs. The regions corresponding to  $\alpha$ -helices 8 and 9 were manually adjusted in order to minimise the number of inserts thus creating one large insert of  $\pm 150$  residues. The majority of the insert was excised, leaving 2 residue overhangs. This insert is subsequently defined as insert 3 (Asp259-Phe408). Finally the regions corresponding to  $\beta$ -sheets 15 and 16 were manually adjusted in order to maximise the alignment of hydrophobic residues. The final alignment used for homology modelling (Fig. 2.5) still only retained modest sequence identity between the template sequences and the target sequences ( $\pm 20\%$ ), despite removal of inserts 2 and 3.

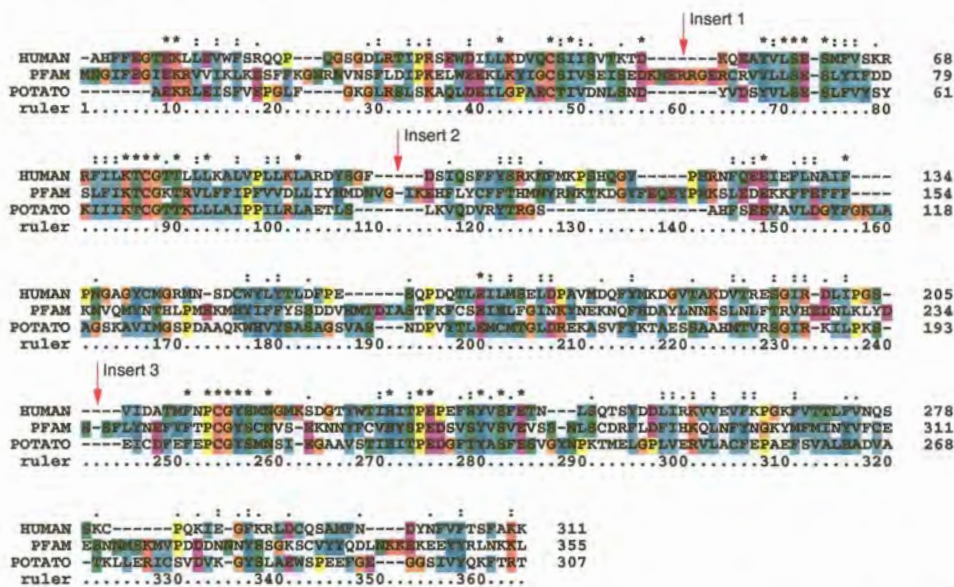


Figure 2.5: Final modelling alignment. The location of *Plasmodium* specific inserts is indicated by the red arrows. Standard CLUSTALX colouring is used (See App. A).

### 2.3.3. Secondary structure prediction

#### 2.3.3.1. Inserts

The secondary structure predictions were first carried out on sequences of the DHFR-TS sequence from various *Plasmodium* species. Attention was focused on those algorithms which could correctly predict the known secondary structure of the *Plasmodium* specific inserts and the junction region from the *P. falciparum* crystal structure. A correct prediction was defined by there being an overlap of predicted structure of four or more contiguous residues with the actual structure. From this, four particularly



successful methods emerged. Success was based on those which could make correct predictions for two or more out of the four known  $\alpha$ -helices. From this grouping those methods which could repeat this feat for three or more out of the four known full *Plasmodium* DHFR-TS sequences were noted. The methods were the GARNIER module from the EMBOSS suite, GOR4, HMMSTER and SAM-T99.

For the bifunctional AdoMetDC/ODC sequence it was sometimes required to split the sequence into smaller overlapping segments for separate submission for certain of the algorithms (JNETPRED, JNETHMM, JNETALIGN, JNETPSSM, JNETFREQ, JPRED and HMMSTER). For the shorter insert 2 there is good agreement between the different methods. The consensus for the short AdoMetDC insert is that it comprises an N-terminal  $\beta$ -sheet and a C-terminal  $\alpha$ -helix, with a coiled region in between (Fig. 2.6).

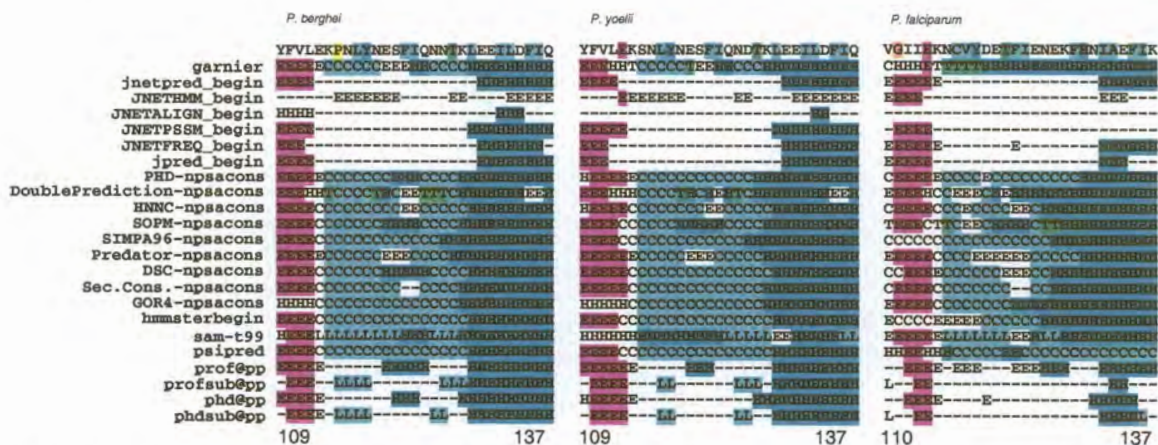


Figure 2.6: Predicted secondary structures of AdoMetDC insert 2. T: turn, C/L: coil, H: helix, E:  $\beta$ -sheet. Image generated with CLUSTALX using standard colours for protein sequences (See App. A).

The predictions for the long insert 3 are less similar. There appears to be a similar distribution and ordering of  $\beta$ -sheets and  $\alpha$ -helices in the beginning of the insert. The rodent inserts are longer, with what appears to be a C-terminal portion that is not present in the human parasite. The conserved regions are shorter in the rodent species. Further details are given in Fig. 2.7.

### 2.3.4. Homology modelling

#### 2.3.4.1. Overall model characteristics

The MODELLER run yielded 2 models with 10 or less residues ( $\pm 3\%$ ) in generously allowed and disallowed regions of the Ramachandran plot, and all of the initial 100 models had 5% or less residues in these regions. Of these two models one which had 5 residues in the generously allowed regions and 1 residue in disallowed regions was chosen for minimisation in INSIGHTII (Fig. 2.8). Minimisation of the chosen model resulted in a slightly poorer Ramachandran plot (RMSD = 1.85 Å with human template), with a decrease in residues found in most favoured regions, from  $\pm 87\%$  before to  $\pm 74\%$  after (Fig. 2.8). The total change in energy was  $3.6 \times 10^5$  kcal.mol<sup>-1</sup> (Final energy: -5664.8 kcal.mol<sup>-1</sup>). A similar increase in RMSD with the potato template (1.64 to 2.22 Å) was observed (Table 2.2). The number of residues found in generously allowed and disallowed regions was constant before and after minimisation. Most residues



*P. falciparum*:

260

SDADKEVTTHIYSTRGTYEDTGMVNCVDVIYKNESTLLNRNNIENIPSIENKESNNNSRCHNNNYSGSCHNIV  
 CCCC-EEEEEECCCCCCCCEEEEEECCCC-CCCCCCCCCCCCCCCCCCCCeeCCCCCCCCEEEE

SVVPSEARNNDHVHRRHYEDTLNRSNISAEDNRRNAQPEKEKEDVRRDDEENKVLIKMIDTNLYECINYNKESF  
 EEE-CCCCCCEee--CCCCCCCCCCCCCCCCCCCC-HHhhhhHH--EEEEEECCHeeEECCCCC-

408

*P. berghei*:

260

SESEKIEVCNSSVYSFDSVENS SKYTSNTANTNITVNVSKDLNNSINTLYNTIDDQTVMLTCSDDKNNLNE  
 CCCCEEEEEECCCEEECCCCCCCCCCCCCCCCEEEEEECCCCCCC-eeEEe-CCCCEEEEECCCCCCCC-H

FDSKLLAEENSKNMDKDNTQVFDNNNYIFISTPKAQETYSNTNSVRSNEKSTCSSNTYTSLLQNDLKEFHFKN  
 HHHHHHHHHH-CCCCCCC-EECCCCEEEECCCCC-CCCCC-CCCCCCCCC-----HHhHHH-CCCC

NSMDTIEDDGKLLVEEEVI SYISNE ETIDAKVEQLDNLNKNMLDNNLDKNYEQTANGPSLSSFTIACS DFLN  
 CCCC--CCC-EEE-----EEeCC-HHHHHHHHHHH-CCCC--CCCCC-----CCCCCCC-EEE---HHHC

KNANDEKGDNTSSNMGIKKIEENLYECINIQNN  
 CCCCCCCCCCCCCC-----HHHHHHHH-----CCC

515

*P. yoelii* :

260

SESEKIEVCSSVYSFDSVENS SKYTNTANTNVTVNVSKDIDNNSINTLYNTVDDQTVILTCSDDKNNLNE  
 CCCCEEEEEE-EEEECCCCCCCCCCCCCCCC-EEEEEECCCCCCC-eEEEE-CCC-EEEEEECCCCCCCCC

SDSKLLLAGENSKNMDKDNTQVFNDDNNNYIFISNPKVQETYSNTNSVRSNEKSTCSSSSTYTSLLQNDLKEF  
 C--Hhhhh-CCCCCCCCC-EECCCCCCCC-CCCC-EE-CCCC-CCCCCCCCCCCC-HHHHHHHHH-

HPKNNAMDTIEDDGRLLVEEEDI SYISNE ETIDAKVEQIDNLSNKSILDNNLDKNYDQNEPSLSSFTVAC SDLI  
 CCCCCC-----CCC-EEEE--C-EEEC-HHHHHHHHHHHCCCC--HCCCCCCCCCCCCCCC-EEEE-hhhh

NKNANDEKSEDNTSKNIRIKKIEENLYECINTQNN  
 CCCCCCCCCCCCCC-----HHHHHHHHH-CCCC

516

Figure 2.7: Consensus secondary structure predictions for AdoMetDC insert 3. Consensus is defined as  $\geq 60\%$  majority coil/sheet/helix amongst all methods which supplied predictions. C: coil, H: Helix, E: Sheet, -: no consensus. e:  $\beta$ -sheet, manual (by eye) assignment (i.e. weak consensus), h: $\alpha$ -helix, manual assignment. Three sets of secondary structural elements can be observed, which are conserved in the rodent species. The first set consists of two  $\beta$ -sheet regions (red). The second set consists of three  $\beta$ -sheet regions (blue). Finally the third set of elements comprises an  $\alpha$ -helical region followed by two  $\beta$ -sheets (green). These sets of elements are separated by predicted coiled regions, which are shorter in the rodent species. Finally those predicted secondary structural elements which are unique to the rodent species are underlined.



found in disallowed regions occurred either in loops or turns, with the exception of Arg249 at the end of helix 8. Helix 8 occurs in a region of low homology and thus may be the result of a misalignment. The root mean square deviation (RMSD) of the backbone  $C_{\alpha}$  atoms between this initial model and templates were 1.45 and 1.65 Å for the human and potato templates, respectively (Table 2.2).

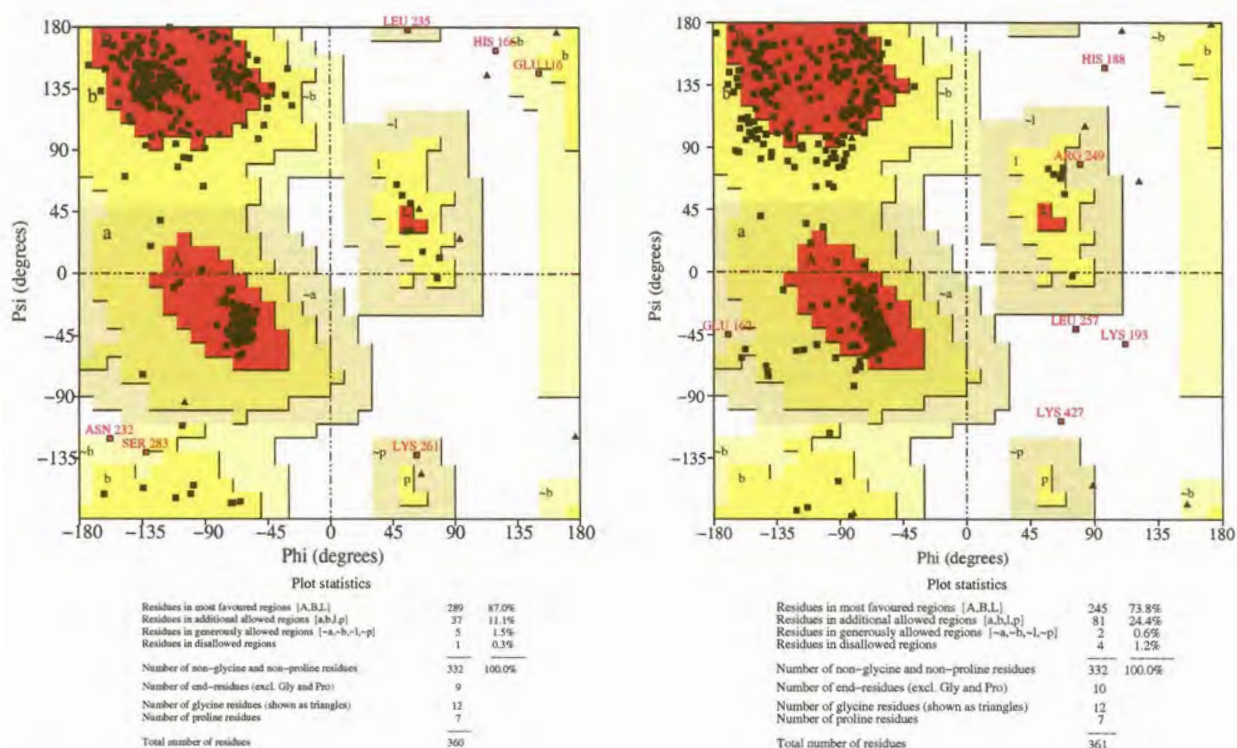


Figure 2.8: Ramachandran plots of the initial and final models. MODELLER model (left), final model after minimisation (right). Most sterically favoured regions (red), additional allowed regions (dark yellow), generously allowed regions (light yellow), disallowed regions (white).  $\alpha$ -helix (A),  $\beta$ -sheet (B), left-handed-helix (L). The presence of MeAdoMet in the final model is reflected by the increased residue count.

Table 2.2: Backbone deviations (Å). In order: Human template, potato template, best initial MODELLER structure, minimised MODELLER structure. Unmutated and mutated models docked with putrescine are also shown (Section 3).

	Human	Potato	MODELLER	Minimised	Unmutated	Mutated
Human	0.0	2.45	1.45	1.85	2.04	1.81
Potato		0.0	1.64	2.22	2.48	2.18
Modeller			0.0	1.4	1.87	1.3
Minimised				0.0	1.28	0.57
Unmutated					0.0	1.38
Mutated						0.0

The topology of the final minimised model is similar to that of the human and potato templates (Fig. 2.9). This model comprises two  $\beta$ -sheets of 8 strands each, and 2 layers of helices. The helix layer lying adjacent to the  $\beta$ -sheet bearing the pyruvoyl residue is made up of 5 helices, while the other helix layer is made up of 6 helices. Secondary structural elements superimposed fairly well (Fig. 2.10). Large deviations were seen in regions where homology was difficult to ascertain, namely in helices 8 and 9, and strands 15 and 16. The more easily identified elements superimpose better. Exceptions to this are strands 10 and 11, which lie adjacent to helices 8 and 9. In general, the  $\alpha$ -helices showed most deviation, while the

$\beta$ -strands near to the active site, showed the best superimposition. The active sites superimposed well for  $\beta$ -strands 5, 12 and 13. A slight shift of 1-1.5 Å for  $\beta$ -strands 3, 11 and 4, as well as the irreversibly bound methyl-ester substrate analogues was observed.

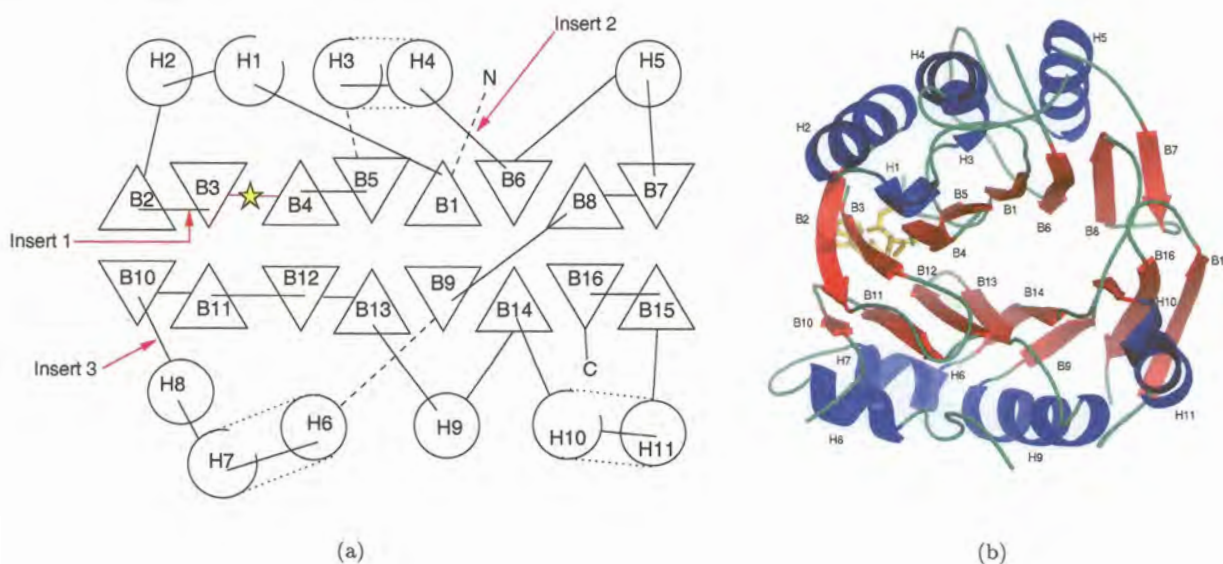


Figure 2.9: Topology of the final model. (a) Topology diagram, closed circles:  $\alpha$ -helices, open circles denote: 3-10-helices, triangles:  $\beta$ -strands, dashed lines: 3-10-helix contiguous with an  $\alpha$ -helix, star: site of proteolytic cleavage/active site. Inserts are as indicated. (b) Cartoon diagram,  $\beta$ -strands: red, helices: blue, MeAdoMet ligand: yellow, loops: green.



Figure 2.10: C- $\alpha$  trace superimposition of model (blue) on human template (green).

#### 2.3.4.2. Active site residues

Within the active site (defined as all residues within 6Å of the ligand) all conserved residues show a similar orientation. Some of the more important residues are shown in Fig. 2.11. Cys82<sub>hum</sub> is a



catalytically important residue in the human enzyme, most likely required for protonation of the  $\alpha$ -carbon of the substrate cysteinyl moiety after decarboxylation (Xiong *et al.*, 1999). The sulphhydryl group of this residue, and that of Cys87 points towards the bound inhibitor, while that of Cys87*pot* is orientated away. A similar observation is made for Glu72*pot* which is more embedded within the active site pocket, while Glu72 and Glu67*hum* lie further out as a result of the presence of the bound inhibitor. His243*hum* has been identified as important for processing and catalysis (Stanley and Pegg, 1991), and has been suggested to take the place of Cys87*hum* to a small degree (Tolbert *et al.*, 2001). The imidazole rings of His243*hum* and His249*pot* both superimpose in the same plane, however, the rings are rotated 180° with respect to each other, which results in a slightly different orientation for the imidazole nitrogens. The model equivalent (His434) has the same orientation as that of the potato enzyme. In all three cases visual inspection reveals that this histidine residue is potentially stabilised by three hydrogen bonds.

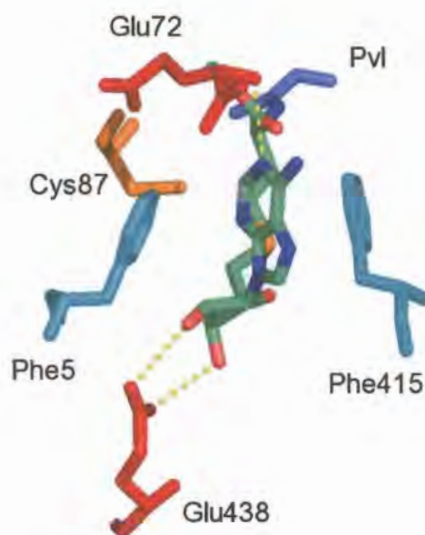


Figure 2.11: Important model-active site residues. Residues are coloured according to their overall nature: acidic residues (red), aromatic (cyan), polar (orange), modified (slate). Pvl: pyruvoyl. The ligand (MeAdoMet) is coloured by atom: N (blue), O (red), C (green), S (orange). Hydrogen bonds are indicated by dashed lines (yellow).

Whereas the malarial enzyme shows considerable divergence from other eukaryotes, there are only three mutations within the active site itself (Table 2.3), and one further mutation revealed by inspection of the surrounding region. Thr245*hum* is replaced by Ser436. While Thr245*hum* makes hydrophobic contact with the inhibitor, no interactions are reported by LIGPLOT for Ser436. Adjacent to this position, Ile242*hum* is replaced by Tyr435 in *P. falciparum*. Near the mouth of the active site Asn224*hum* is replaced by Thr416 in *Plasmodium*, and Gly3 replaces His5*hum*. The histidine residue is about 2 Å closer to MeAdoMet (Fig. 2.12).

A number of important interactions observed between the human enzyme and the substrate are repeated in the model docked with MeAdoMet. The adenine ring of the MeAdoMet is stacked via



Table 2.3: Comparison of model and human active site residues. Included residues are based on LIGPLOT results, previous mutation studies and visual inspection. Parenthesis indicate that no contact was predicted by LIGPLOT. Substitutions between species are in italics.

Model residue	Human residue	Model residue	Human residue	Model residue	Human residue
( <i>Gly3</i> )	( <i>His5</i> )	Ser74	Ser69	Tyr420	Tyr228
Phe5	Phe7	(Thr86)	Thr81	Ser421	Ser229
(Glu6)	(Glu8)	Cys87	Cys82	(His434)	(His243)
(Glu9)	(Glu11)	Thr90	Thr85	<i>Tyr435</i>	<i>Ile244</i>
Leu70	Leu65	Phe415	Phe223	Pro437	Pro246
Ser71	Ser66	<i>Thr416</i>	<i>Asn224</i>	Glu438	Glu247
Glu72	Glu67	Cys418	Cys226	( <i>Ser436</i> )	<i>Thr245</i>
Pvl73	Pvl68	Gly419	Gly227		

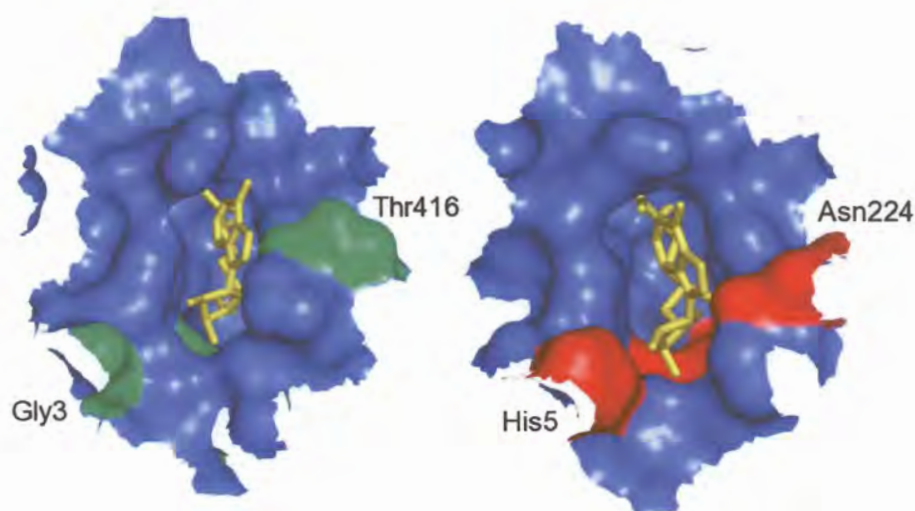


Figure 2.12: Substitutions in the model active site. Van der Waals surfaces of the model (left) and human (right) active sites are shown. Substituted residues are indicated in green and red. Tyr435, Ser445, Ile224<sub>hum</sub> and Thr245<sub>hum</sub> (unlabelled) are visible behind the MeAdoMet ligands (yellow). Surfaces generated with PYMOL.

hydrophobic interactions between two phenyl rings (Phe5 and Phe415), as in the human structure (Phe7 and Phe223). Mutation studies in the human enzyme suggest these residues are important for substrate and inhibitor binding (Tolbert *et al.*, 2001). Glu438 is shown by LIGPLOT to be capable of forming two hydrogen bonds with both hydroxyls of the ribose moiety, as does Glu247*hum* (Fig. 2.13). The distance between the parent and donor atoms was larger in the model (0.33-0.45Å). In both the model and the human structure there exists a hydrogen bond between N1 of the adenine ring and the amide nitrogen of corresponding Glu residues (Glu67*hum*, Glu72). The orientation of MeAdoMet within the active site was fairly similar for both the model and human structure. For the model, the pyruvoyl residue demonstrated more out-of-plane distortion than the template structures. It is expected that the four atoms of the two carbonyl moieties of pyruvoyl will remain in plane in order for pyruvoyl to act as an electron sink as described (Section 1.3.2). The sulphonium-methyl group also displayed a slightly different orientation in each of the structures. In the human crystal structure it points towards Tyr193*hum* of  $\beta$ -strand 11, while in the model it points between Tyr435 of  $\beta$ -strand 13.

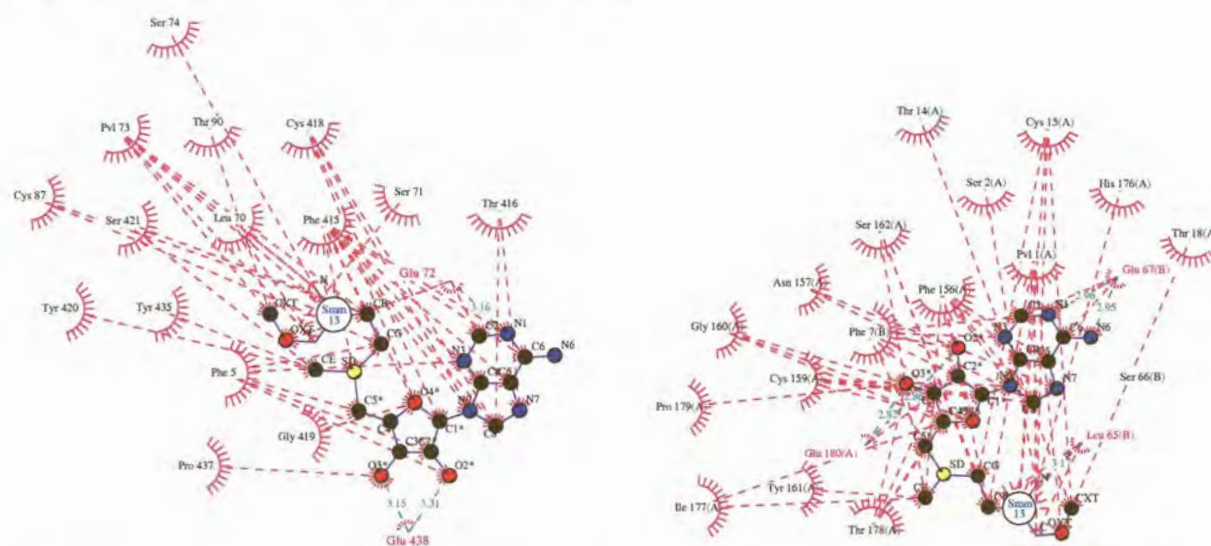


Figure 2.13: Ligand interactions for the model (left) and human (right) active sites with MeAdoMet. “Hedgehogs” indicate hydrophobic interactions. Hydrogen bonds are indicated by green dashed lines.



### 2.3.4.3. Active site shape

The active site forms a well defined cavity, with some noticeable differences between the model and the human template. In the model active site a cavity can be distinguished in the vicinity of the pyruvoyl carbonyl oxygen when generating surfaces in PYMOL with hydrogens removed, or using a Conolly surface generated with a 1.3Å probe (INSIGHTII). This cavity is centred around Asn423, and surrounded by residues Ser421, His434, Ser74, Leu69 and Leu70. No such cavity is observed in the human structure around the corresponding Asn231<sub>hum</sub> residue (Fig. 2.14). Another cavity which can be clearly seen in the human enzyme, is however, only visible in the model if surfaces are generated with hydrogens removed. This cavity is centred around Tyr252<sub>hum</sub>, and is surrounded by Gly9<sub>hum</sub>, Glu11<sub>hum</sub>, Ser254<sub>hum</sub>, His243<sub>hum</sub>, Ile244<sub>hum</sub> and Thr245<sub>hum</sub>. In both species these cavities are located near the sulphonium methyl group of MeAdoMet (Fig. 2.15), and both cavities are partly described by a tyrosine (Tyr252<sub>hum</sub>, Tyr443) and glutamate residue (Glu11<sub>hum</sub>, Glu9). The glutamate residue is conserved across all species, and situated between 8 and 9 Å from the sulphonium atom.

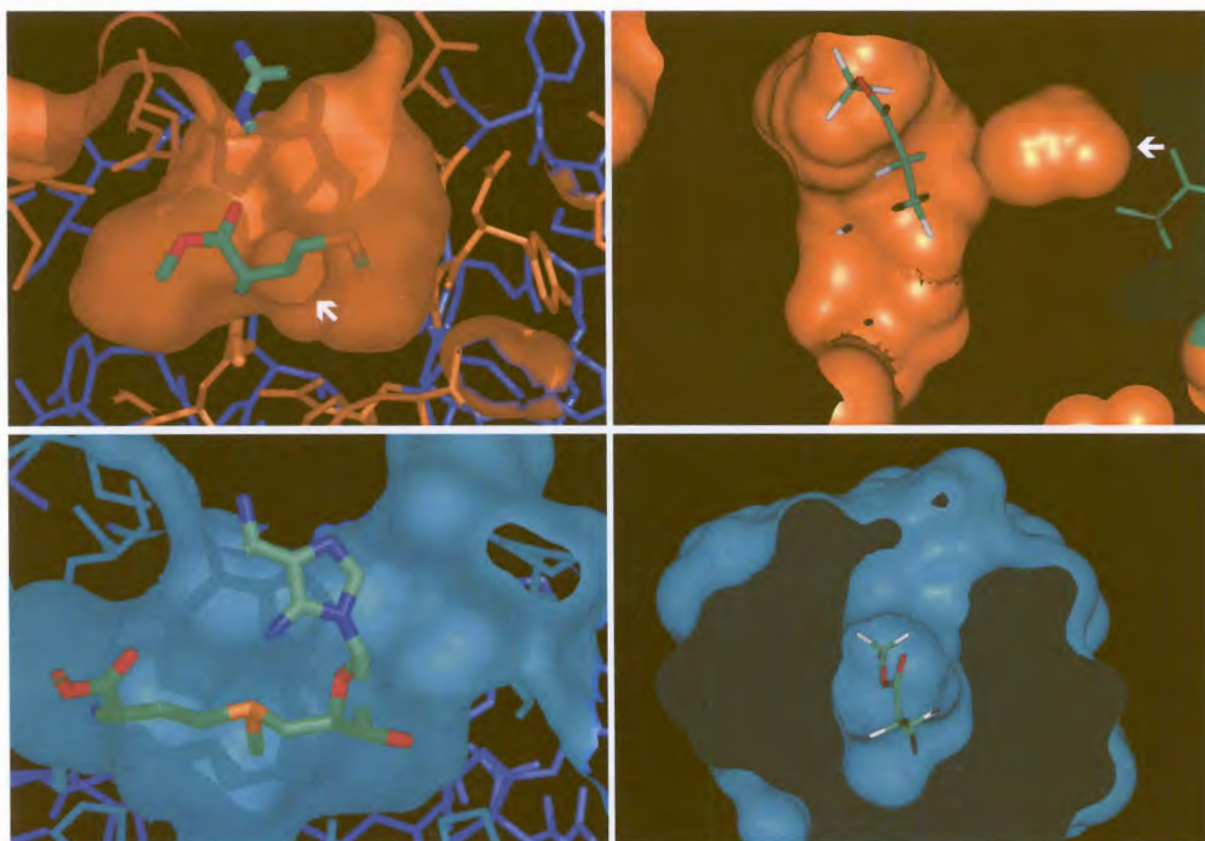


Figure 2.14: Active site shapes of AdoMetDC. Top: model active site. Bottom: human active site. The MeAdoMet ligand is included: C (green), N (blue), S (orange), O (red). Surfaces generated in INSIGHTII (right) and PYMOL (left). White arrows indicate extra cavities.



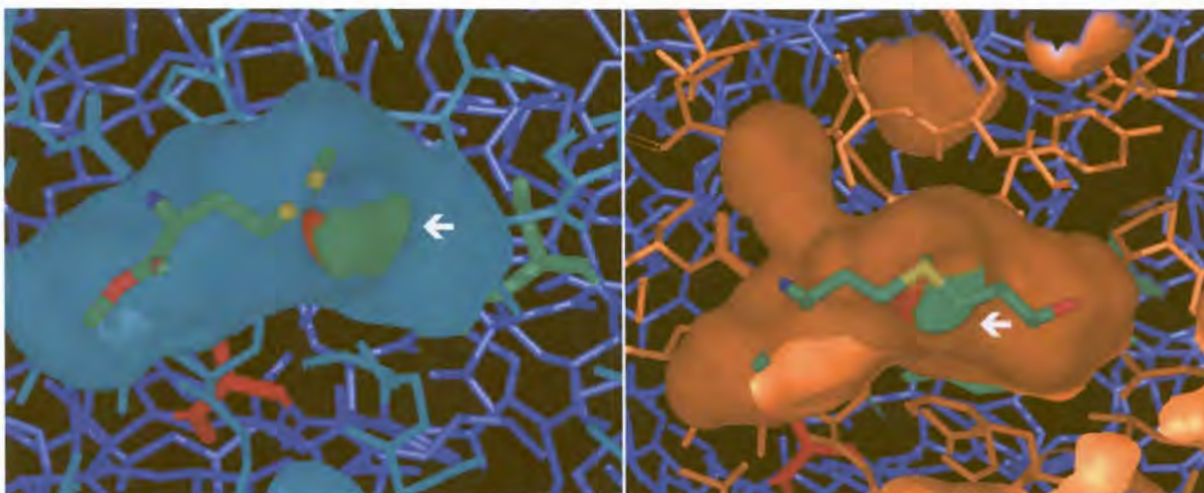


Figure 2.15: Cavity identified near the sulphonium group (white arrow). Human (left) and model (right) structures. C (green), S (yellow), N (blue), O (red). Surfaces generated with PYMOL.

#### 2.3.4.4. Structure of insert 1

Insert 1 was modelled *ab initio* as a random coil. Further analysis of this loop suggests that despite the divergence of malarial AdoMetDC from other eukaryotic sequences, structure may be retained through complementary mutations. This is seen in the form of a charge network, which resides between helix 1 and insert 1. Helix 1 is a 3-10 helix comprising a DLRT motif in humans, GLRS in potato, and SFLD in the model (Fig. 2.16). In the human structure, the sidechains of Asp31<sub>hum</sub> and Arg33<sub>hum</sub> from helix 1 are within electrostatic bonding distance of each other. Arg33<sub>hum</sub> is also situated close to Asp58<sub>hum</sub> and Asp60<sub>hum</sub>. Furthermore, Glu61<sub>hum</sub> is located adjacent to Lys56<sub>hum</sub> in  $\beta$ -strand 2. In the potato structure this network is not as extensive, with Arg38<sub>pot</sub> (equivalent of Arg33<sub>hum</sub>) being located near to Asp63<sub>pot</sub> (Asp58<sub>hum</sub>) and Asp66<sub>pot</sub> (Glu61<sub>hum</sub>). The corresponding potato residues for Asp31<sub>hum</sub> and Lys56<sub>hum</sub> are not similarly charged (Gly and Tyr, respectively). A more extensive charge network is seen in the model. Arg38<sub>pot</sub>/Arg33<sub>hum</sub> is replaced by Leu31. Glu61<sub>hum</sub>/Asp66<sub>pot</sub> and Asp58<sub>hum</sub>/Asp63<sub>pot</sub> are in turn replaced by Arg66 and Arg64, respectively. Asp56 of insert 1 is favourably located in the opposite arm of the loop to interact electrostatically with these Arg residues. Asp52 (Lys56<sub>hum</sub>) may also possibly interact with Arg66, but this residue is possibly too far away ( $\pm 11$  Å), and its sidechain is not favourably orientated in the model. More residues further out in the loop may contribute to these interactions: Lys57, Asp59 and Arg61. However, these residues are replaced by hydrophobic residues in the alignment with *P. knowlesi*, *P. yoelii* and *P. chabaudi*.



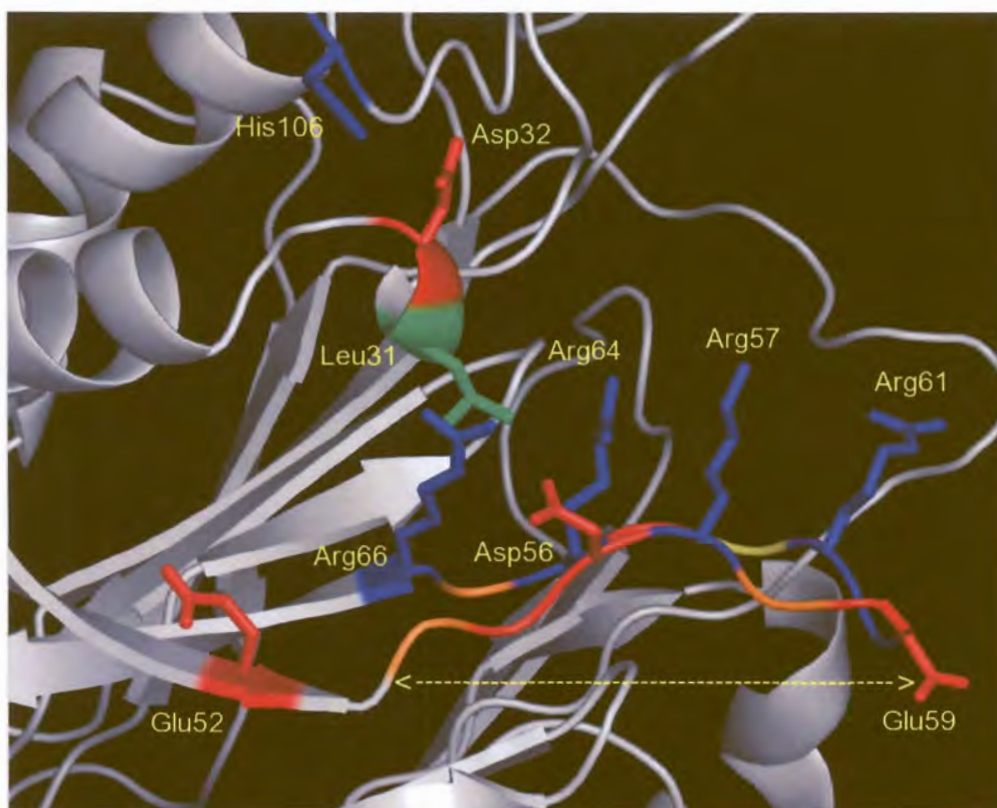


Figure 2.16: Charge network associated with insert 1. Acidic (red), basic (blue), hydrophobic (green) residues referred to in section 2.3.4.4 are displayed as stick models. Insert 1 is indicated by the arrows.

## 2.4. Discussion

### 2.4.1. Identification of other *Plasmodium* sequences

The bifunctional arrangement of enzyme activities is not unusual in *Plasmodium*, among them being DHFR-TS (Yuvaniyama *et al.*, 2003), glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase (Glc6PD-6PGL, Clarke *et al.*, 2001) and dihydrofolate synthetase-folylpolyglutamate synthetase (DHFS-FPGS, Salcedo *et al.* 2001). In some cases, such as for DHFR-TS and Glc6PD-6PGL, the two activities catalyse consecutive steps of a metabolic pathway. The bifunctional arrangement may therefore allow substrates to be efficiently channelled from one enzyme to the next, as has been suggested for DHFR-TS (Yuvaniyama *et al.*, 2003). Substrate channelling is not an option for malarial AdoMetDC/ODC, because these activities are not consecutive in a metabolic pathway (Müller *et al.*, 2000). Both activities are rate-limiting steps of polyamine metabolism (Marton and Pegg, 1995), and it has been suggested that co-expression of these activities allows efficient regulation of polyamine biosynthesis (Wrenger *et al.*, 2001). During this study complete bifunctional sequences were identified for the rodent malaria parasites *P. berghei* and *P. yoelii*. Fragments were also identified for *P. chabaudi* and the primate parasite *P. knowlesi*. All the bifunctional AdoMetDC/ODC sequences are similar in length ( $\pm 1400$  amino acids) and display a similar pattern of open reading frames for the surrounding genomic regions. The radiation of *Plasmodium* is thought to have occurred between 55 and 129 million years ago. Divergence of the



vertebrates has been set at approximately 200 myr ago (Ayala *et al.*, 1998). No other bifunctional AdoMetDC/ODC enzymes have been discovered for organisms other than *Plasmodium* (Müller *et al.*, 2000; Wrenger *et al.*, 2001). It therefore appears that these genes are cognate across the *Plasmodium* species for which there are currently sufficient sequence data. Since the gene appears to exist in divergent malarial species it is suggested that all *Plasmodium* species have the bifunctional AdoMetDC/ODC enzyme. Thus it seems that bifunctional ODC/AdoMetDC may have evolved only once, early in the *Plasmodium* lineage.

## 2.4.2. Sequence properties of the *Plasmodium* AdoMetDC/ODC

### 2.4.2.1. Conservation of secondary structural elements

Initial attempts to model *P. falciparum* AdoMetDC were hindered by the low homology of the AdoMetDC domain of the bifunctional enzyme to the template enzymes and the presence of *P. falciparum*-specific amino acid inserts. This prompted the search for other *Plasmodium* sequences. Including these sequences made it easier to identify secondary structural elements of the template structures which were highly diverged in the *P. falciparum* sequence, and to identify with greater certainty where the *Plasmodium*-specific inserts lie. When comparing the *P. falciparum* protein sequence with the template structures, the more conserved elements were seen to cluster around the active site, while more diverged elements congregated on one side of the protein (Fig. 2.4). The alignment of template secondary structures was also largely in agreement with secondary structure predictions of the target sequence. Even with this extra information, it was difficult to identify the corresponding regions of helix 8 and helix 9 from the human template. According to the model these elements are located in the N-terminal region of insert 3 of the AdoMetDC domain. This insert is the most variable of the *Plasmodium* specific regions, with those of the *P. yoelii* and *P. berghei* sequences being even longer ( $\pm 100$  extra residues). Because this region is highly variable it is considered possible that these helices have been incorrectly identified. In order to resolve this, more experimental evidence is needed, specifically, exact delineation of the *Plasmodium* specific inserts.

The remaining secondary structural elements of the human enzyme which gave difficulty ( $\beta$ -strands 7, 15 and 16), occupy the region responsible for dimerisation within the human enzyme. Furthermore,  $\beta$ -strands 14 and 15 are found in the C-terminus of the *P. falciparum* AdoMetDC domain which leads to the hinge that connects the two domains. For *Plasmodium* it is considered likely that similar regions are responsible for interaction between the AdoMetDC domains within the heterotetrameric complex, since this would represent the most parsimonious evolutionary route. Most of the interactions responsible for making up the bifunctional complex are also expected to originate from the hinge region. This has been partially confirmed using recombinantly expressed *P. falciparum* ODC. *P. falciparum* ODC which lacks the hinge region has lower activity than recombinant ODC(+hinge) (Birkholtz *et al.*, 2004). ODC is an obligate dimer (Seely *et al.*, 1982; Almrud *et al.*, 2000), and therefore the lowered activity may



be due to loss of interactions between the two ODC monomers. Furthermore, the ODC(+hinge) has been demonstrated to associate with separately expressed AdoMetDC, an interaction which is lost with ODC(-hinge) (Birkholtz *et al.*, 2004). Because  $\beta$ -strands 14 and 15 correspond with the dimerisation interface of the human enzyme, and are in close linear proximity with the hinge region, it is expected that these elements will therefore also be involved in domain-domain interactions. In effect these regions are proposed to take on additional functions when compared to the human enzyme, namely, assisting the association of the bifunctional complex. The evolution of additional functions for these regions represents a transition from a more aqueous environment for monofunctional AdoMetDC to a more protein rich environment in the bifunctional complex. Due to this, it is expected that these regions will be highly diverged. When considered together, the more diverged elements are seen to cluster on one face of the model, namely the  $\alpha\beta$ -slice of the  $\alpha\beta\beta\alpha$  sandwich which does not contain the pyruvoyl moiety (Fig. 2.4). For similar reasons, it is therefore predicted that this region as a whole makes physical contact with the rest of the bifunctional complex.

#### 2.4.2.2. *Plasmodium*-specific inserts

The inserts display a number of interesting properties. The shorter inserts are much more conserved than the longer inserts. Insert 1 is generally conserved between *P. berghei*, *P. yoelii* and *P. knowlesi*, however, that of *P. falciparum* contains more charged residues. Insert 2 shows a similar distribution of hydrophobic and polar residues between the *Plasmodium* species (Fig. A.1, App. A). Insert 2 is only one residue shorter in *P. falciparum* and in the *P. knowlesi* fragment. Secondary structure predictions for insert 2 from all of the complete bifunctional sequences using various algorithms give similar results (Fig. 2.6). The consensus is that insert 2 contains a  $\beta$ -strand in the N-terminal region and an  $\alpha$ -helix in the C-terminal region, separated by a random coil. The adjacent secondary structural elements in the core structure which was modelled are of opposite types in each case (i.e.  $\alpha$ -helix :  $\beta$ -sheet), thus these regions of predicted secondary structure within the inserts are unlikely to be extensions of that found in the core structure. Insert 2 therefore probably represents a distinct structural region unique to *Plasmodium* AdoMetDC. The shorter AdoMetDC inserts occur on the same  $\alpha\beta$ -slice of the protein carrying the pyruvoyl residue. Due to the greater conservation of these inserts between the *Plasmodium* species and the greater divergence of the other  $\alpha\beta$ -slice, they are expected to be more structurally important for the AdoMetDC domain than for the rest of the bifunctional complex.

Insert 3 shows considerable variation between the *Plasmodium* species compared to the shorter inserts. Firstly, the rodent inserts are considerably longer ( $\pm 100$  residues), and secondly, more divergence is visually detectable on the sequence level. Greater similarity can be found when comparing the secondary structure predictions. When analysed on this level, the rodent sequences appear compressed, with shorter coiled regions connecting the secondary structural elements that are shared between the three species. Furthermore, insert 3 of each of the rodent species is extended in the C-terminal direction. The greater C-terminal variation partly justifies the manual adjustments that were made in order to align helices 8



and 9 from the human enzyme. It was decided that these elements are more likely to lie before insert 3 due to slightly greater sequence conservation between the template and target sequence. It is predicted that the more conserved regions play a greater role in the native functioning of AdoMetDC/ODC, and as a result the C-terminal extensions may possibly be dispensable. This insert is also seen to reside on the  $\alpha\beta$ -slice that is more diverged, and may therefore mediate bifunctional complex formation. However, it is located on the opposite side to where dimerisation between the AdoMetDC domains is predicted to occur. Deletion mutagenesis of a region encompassing this insert has revealed that it may be required for AdoMetDC activity, but not for bifunctional complex formation (Birkholtz *et al.*, 2004). However, this study was conducted before anything about the structure of *P. falciparum* AdoMetDC was known, and the deletion was based on a previously published alignment (Müller *et al.*, 2000). Based on this alignment, *P. falciparum* AdoMetDC was predicted to have only one large insert. This study reveals that three inserts are more likely. As a consequence, this deletion also contained portions of what is now predicted to be the core structure of AdoMetDC. Deleting these core regions are expected to have a profound effect on the normal functioning of *P. falciparum* AdoMetDC. Therefore, further studies focusing on the new predicted insert are required in order to determine its role.

Whereas the shorter insert 2 is expected to have a more structural role, this is less certain for the longer insert 3. Long inserts are common in *Plasmodium* and often occur in genes encoding important housekeeping functions (Pizzi and Frontali, 2001). The function of these inserts is unknown, however, they have been demonstrated via deletion mutagenesis to possess functions specific to the enzymes they occur in, for example in malarial DHFR-TS (Yuvaniyama *et al.*, 2003) and AdoMetDC/ODC (Birkholtz *et al.*, 2004). These inserts are often characterised by low complexity, i.e. biased amino acid composition (Pizzi and Frontali, 2001). The *Plasmodium* genome itself is extremely biased in composition as witnessed by its A+T-richness (80%, Gardner *et al.*, 2002). It would therefore appear that there has been considerable evolutionary pressure on the *Plasmodium* genome. It has been suggested that this bias can manifest itself on the protein level in the form of low amino acid sequence complexity, and that amino acid bias may therefore not be as the result of evolutionary pressure on the proteins themselves (Singer and Hickey, 2000). Xue and Forsdyke (2003) have also suggested that these inserts may represent potential folding regions on the DNA level, much as is the case for introns. There is however, little evidence that these *Plasmodium* specific inserts are interchangeable with introns. Insert 3 does not possess the NND and NNI repeats that are seen in many other *Plasmodium* inserts, it does however have an Asn and Lys bias. Based on current knowledge it is therefore suggested that insert 3 may have a functional role in AdoMetDC/ODC, but that at some point genome bias played a role in its evolution.



### 2.4.3. Homology modelling

#### 2.4.3.1. Overall model characteristics

The model displays good properties considering the large degree of sequence divergence between the target and template sequences ( $\pm 20\%$  after removal of inserts 2 and 3). The RMSD of the C- $\alpha$  atoms between the target and templates was 1.85 Å and 2.22 Å (Table 2.2) for the human and potato target templates, respectively. This was even better than that between the templates themselves (2.25 Å) which have a higher sequence identity of  $\pm 30\%$ . Deviations as large as 1.5 Å can be expected for target-template sequence identities between 50-90% (Krieger *et al.*, 2003). It was demonstrated that most residues within a protein are restricted with regards the  $\phi$  and  $\psi$  angles of the peptide bond due to steric clashes introduced by the sidechains (Ramachandran *et al.*, 1963). This was also observed for the model: in the initial model and final minimised model over 90% of the residues were in highly favoured regions (Fig. 2.8). Longer minimisations tended to make this plot worse, with more residues moving into the less favoured region. This was also observed for the human template (results not shown) and therefore considered to be an artefact of the energy function, and not the model itself. It is known that once such energy functions have removed the worst violations there tends to be an accumulation of small errors (Krieger *et al.*, 2003). It was therefore decided to modify the default parameters and use a minimisation protocol that terminated relatively quickly ( $\pm 1200$  steps) in order to limit this. Although the combined number of residues in generously allowed and disallowed regions remained the same, the number in completely disallowed regions increased. The offending residues mostly occur in loop or turn regions however, where greater flexibility is expected. The exception is Arg249 which occurs at the C-terminus of helix 8 in the model. Helix 8 was possibly incorrectly identified due to its proximity to AdoMetDC insert 3 (Section 2.4.2.2), thus this violation may therefore be the result of misalignment. Further concerns are also raised due the fact that the sidechain of Arg249 is partially buried. The guanidinium moiety however, is exposed to the surface, and thus the positive charge is possibly neutralised by water. Visual inspection also reveals that the sidechain of Asp253 is near enough to possibly neutralise the positive charge. The secondary structural elements that superimposed best with the template cognates tended to be those that were more conserved.  $\beta$ -strands 10 and 11 deviated from this trend. These strands sit adjacent to the highly diverged helices 8 and 9. This divergence is expected to affect those regions in immediate contact and therefore this deviation is not unexpected. For most of the elements surrounding the active site there was good superimposition, indicating that elements were correctly placed for the proper functioning of AdoMetDC. Slight differences described next are observed however, indicating that this enzyme is exploitable for rational inhibitor identification (Section 2.4.3.2).

The final model falls into the same  $\alpha\beta\beta\alpha$ -fold as for human and potato AdoMetDCs (Ekstrom *et al.*, 1999; Bennett *et al.*, 2002), and has overall a similar topology to the human and potato templates (Fig 2.9). The  $\beta$ -sheets are comprised of the same number of  $\beta$ -strands in the model and templates, indicating these regions were correctly aligned. There are differences in the number of 3-10-helices and  $\alpha$ -helices,



however. The 3-10-helix 5 from the human structure has no counterpart in the malarial model and is modelled as a random coil. There are two possible reasons for this. Firstly, the helix motif in the human structure is PSHQG whereas the malarial motif is KTKDG. Since proline has an atypical structure compared to the other 19 amino acids it tends to constrain flexibility, and it is not surprising that replacement of this residue results in an altered and possibly more flexible structure. Furthermore, the corresponding region in the potato template is unresolved and thus could not contribute to the modelling of this region. An extra helical region is also present in the non-pyruvoyl  $\alpha\beta$ -slice of the model that consists of a 3-10 helix that is contiguous with an  $\alpha$ -helix (helices 10 and 11). This can be accounted for due to the fact that this region is unresolved in the human template, but is revealed to be helical in the potato template. Due to uncertainties regarding the interacting regions within the bifunctional complex, and the current methods for modelling of protein-protein interactions, no dimerisation of *P. falciparum* AdoMetDC was attempted *in silico*.

#### 2.4.3.2. Active site composition

All of the residues that have been previously identified as being important for correct enzyme function show similar orientations to the human template (Table 2.3). A few differences are observed when comparing residues of the model with the potato template. Firstly, the sulphhydryl group of Cys87 points towards the bound inhibitor, while that of Cys87<sub>pot</sub> is orientated away. Secondly, Glu72<sub>pot</sub> is more embedded in the active site pocket, while Glu72 and Glu67<sub>hum</sub> lie further out. These differences are introduced by the irreversibly bound inhibitor which is present in the human template, and is modelled in the malarial structure. The Glu72<sub>pot</sub> occupies a region that is occupied by the bound inhibitor in the other two structures. The Cys87 equivalent in the human is required for both processing and activity, therefore it is important that this residue is in the correct orientation. Enzymes frequently undergo structural shifts on binding of substrates and inhibitors and this can be used to explain these differences. A third difference is observed for His434. Due to rotation of the imidazole ring between the human and potato templates, the imidazole nitrogens occupy opposite ends of the imidazole plane in each structure. The model residue has the same orientation as that of the potato enzyme. This difference in the templates may simply be a result of the difficulty in distinguishing between the electron density of carbon and nitrogen atoms (Davis *et al.*, 2003). In all three cases inspection reveals that this histidine residue is potentially stabilised by three hydrogen bonds. Hydrogen bond networks have been identified as possibly being important for the stabilisation of this residue in the human structure in order to prevent incorrect protonation in the Cys82Ala mutant (Tolbert *et al.*, 2001). Further analysis of the model indicates that in either orientation a number of hydrogen bonds are possible. In the modelled orientation the position of  $N^{\delta 1}$  within His434 is able to form a hydrogen bond with the hydroxyl of Ser445. The  $N^{\epsilon 2}$  His434 can also potentially form a hydrogen bond with Cys432 and Ser74. In the opposite orientation the  $N^{\delta 1}$  atom could hydrogen bond with Ser421 and  $N^{\epsilon 2}$  could form bonds with Ser74 and Cys432, similar to the human template. Thus in either orientation this residue can be stabilised by a similar degree of hydrogen



bonding. Assuming that stabilisation of this residue is also important to prevent incorrect protonation in *P. falciparum* AdoMetDC, the orientations may be equally valid.

The *P. falciparum* model shows similar but less binding interactions with the methyl-ester substrate analogue (Fig. 2.13). The sequence divergence of the active site is less pronounced than for the rest of the enzyme, indicating that most of the active site residues play important roles for correct enzyme functioning. Of the 19 residues predicted to make contact with MeAdoMet in the human structure, three substitutions are observed in the model (Table 2.3). Of the correspondingly positioned model residues, 17 are predicted to contact the ligand, 16 of these being conserved in identity and structure. Thr245<sub>hum</sub> is replaced by the less bulky Ser436 and may therefore explain why this residue does not contact the ligand. Adjacent to this position, Ile242<sub>hum</sub> is replaced by Tyr435 in *P. falciparum*. Since only the backbone atoms of these residues make contact with the ligand and the sidechain points into the protein core, this position may be more susceptible to mutation. Thirdly, Thr416 on the surface of the enzyme takes the place of Asn224<sub>hum</sub>. The last observable substitution, which is not predicted to make contact in either humans or the model, is that of Gly3 replacing His5<sub>hum</sub>. This substitution may explain why Tris is observed to inhibit the human enzyme (Pegg and Williams-Ashman, 1969) but have no effect on the malaria enzyme (Wrenger *et al.*, 2001). Tris has been resolved in various crystal structures of the human enzyme and it has been suggested that His5<sub>hum</sub> forms weak hydrogen bonds with the molecule (Ekstrom *et al.*, 2001; Tolbert *et al.*, 2003b). The corresponding model glycine residue backbone atoms are too far away, and there is no sidechain to interact with Tris. This interaction within the human enzyme may therefore be more important than originally suggested for Tris inhibition (Ekstrom *et al.*, 2001; Tolbert *et al.*, 2003b). Although the substituted residues are few, they are near enough to the ligand that these differences may possibly be exploitable for inhibition. It may be possible to design inhibitors that selectively interact with these differences in the model active site in a manner that does not favour the corresponding interactions in the human binding site.

The important interactions observed between the human enzyme and the substrate are maintained in the model. The adenine ring of the MeAdoMet is stacked via hydrophobic interactions between the phenyl rings of Phe5 and Phe415. Mutation studies in the human enzyme suggest that the corresponding residues are important for substrate and inhibitor binding (Tolbert *et al.*, 2001). Glu438 is capable of forming two hydrogen bonds with both hydroxyls of the ribose moiety (Fig. 2.13), as does the human cognate. In both the model and the human structure there exists a hydrogen bond between N<sup>1</sup> of the adenine ring and the amide nitrogen of corresponding Glu residues of the templates. The orientation of MeAdoMet within the active site was similar for both the model and human structure. For the model, the pyruvoyl residue demonstrated more out-of-plane distortion. It is expected that this residue will remain in plane, in order for it to act as an electron sink as described (Section 1.3.2). However, once bound, the irreversible ligand MeAdoMet could allow for a relaxation of this requirement because it is unable to allow the reaction to run to completion. The sulphonium-methyl moiety also displayed a slightly different orientation in the human template and model. Because the most important contacts are maintained, the



model was considered accurate enough for initial attempts of virtual inhibitor screening. Furthermore, the accumulation of small differences described indicates that novel binders to malarial AdoMetDC could be found.

#### 2.4.3.3. Active site shape

The active site cavity of the malarial AdoMetDC model displays some interesting differences when compared with the human host model. Firstly, in the model an extra cavity can be distinguished in the vicinity of the pyruvoyl carbonyl oxygen. This cavity is absent in the human structure. However, this cavity is only visible when generating surfaces without hydrogens, or when generating a solvent accessible surface using a probe smaller than the standard water molecule (1.4 Å). It is nonetheless considered possible that ligand moieties could be accommodated within this cavity, allowing for some structural shifts within the protein. Another cavity which can be clearly seen in the human enzyme, is however, only visible in the model if surfaces are generated with hydrogens removed. In both species these cavities are located near the sulphonium methyl group of MeAdoMet (Fig. 2.15), and both cavities are partly described by a tyrosine (Tyr252*hum*, Tyr443) and with negative glutamate residues (Glu11*hum*, Glu9) lying adjacent. The glutamate residue is conserved across all species, and situated between 8 and 9 Å from the sulphonium atom in the known structures and the model. It has been noted that potential inhibitors which lack a positive charge that simulates the S<sup>+</sup> atom are less effective (Pankaskie and Abdel-Monem, 1980; Pegg and Jacobs, 1983). The presumably negatively charged residues that may be required for this have yet to be identified (Tolbert *et al.*, 2001). The other negative residues (Glu67 and Glu247) of the human enzyme have already been identified as making hydrogen bonds with the substrate analogue—thus neutralising their negative charge—and are therefore less likely candidates for interacting with the positive group. This leaves Glu11*hum* as the only other residue that could interact with the sulphonium atom. This residue is also required for processing (Table 1.1). Even the conservative mutation of Glu11Asp results in inhibition of human AdoMetDC processing by putrescine (Xiong *et al.*, 1999). Therefore, it is difficult to test this hypothesis using conventional replacement mutagenesis. This cavity in the human enzyme is occupied by two water molecules which may possibly mediate the interaction between the sulphonium atom and Glu11*hum*. The model was generated without water molecules, and this may therefore explain the smaller size of this cavity in the model. It remains uncertain as to whether the water molecules should be included for modelling in these positions, since they are not conserved in the potato structure. There are therefore a number of indications that the active sites of the host and parasite enzyme are sufficiently different to make this enzyme worth exploring for therapeutic intervention.

#### 2.4.3.4. Structure of insert 1

Inserted regions cannot be modelled based on the alignment with the templates because these regions are by definition unaligned. Consequently loop modelling has to either resort to searching databases of known protein structures for similar regions or *ab initio* modelling from first principles. In either case the



current practical upper limit for loop modelling is considered to be in the order of 12 residues (Fiser *et al.*, 2000). Based on this, the two larger inserts (inserts 2 and 3) of the AdoMetDC were considered too large for *ab initio* modelling and were therefore left unmodelled. The MODELLER program was however, able to derive a structure for insert 1 which is only 7 residues long. The resulting structure does not possess any of the typical secondary structures, but is rather modelled as a random coil. According to the model, insert 1 is a surface exposed loop, and the domination of polar and charged residues would suggest there are considerable interactions between this loop and the solvent. Insert regions of *P. falciparum* DHFR-TS play defined structural roles for enzyme functioning (Yuvaniyama *et al.*, 2003). Deletion of the shorter *P. falciparum* insert from ODC results in a less active bifunctional enzyme and mediates protein-protein interactions (Birkholtz *et al.*, 2004). It is therefore still possible that insert 1 of AdoMetDC has an important defined structure and may not be flexible.

As mentioned (Section 2.4.3.1) divergence increases significantly in regions further from the active site. Further analysis of insert 1 reveals that this divergence can be accommodated through mutations complementary in their physicochemical properties (Section 2.3.4.4). The *Plasmodium* charge network could possibly have been improved by aligning LRT:*hum*/LRS:*pot* with LD-:*pfam*. This would have placed Asp32 in the position occupied by Leu31, and made favourable interactions with Arg64 and Arg66 possible. In the current model the side chain of Asp32 is about 10 Å from Arg64 and Arg66. This realignment may also have brought Leu37 of helix 2 into closer contact with helix 4, in order to enable hydrophobic packing. The loop connecting helix 1 with helix 2 would, however, have been shortened from two residues to one. Furthermore, Ile33 would have been moved from a position of favourable hydrophobic burying between helix 1, helix 2 and helix 4, to a possibly surface-exposed position. Based on these considerations, however, and inspection of the multiple alignment including other *Plasmodium* sequences, the original alignment for helix 1 was retained. Asp32 in its current position within the model may nonetheless interact favourably with His106 which replaces Tyr101*hum* and Leu106*pot*. As this analysis is relatively tedious it was not carried out for the entire model structure, therefore the possibility of global complementary mutagenesis still needs to be confirmed. This nonetheless demonstrates that although the *Plasmodium sp.* sequences have undergone considerable divergence, and that complementary mutations may enable the enzyme to retain its function.

A homology model of malarial AdoMetDC was successfully constructed based on the X-ray crystal structures of the human and potato enzymes. Despite the considerable difficulties introduced by the low sequence identity, its quality was deemed adequate to begin initial biochemical investigations based on model predictions. The following chapter describes modelling and mutational analysis that was conducted in order to test some of the predictions made from the model regarding the lack of putrescine stimulation in malarial AdoMetDC.



## Chapter 3

# Model guided mutational analysis of malarial

## AdoMetDC

### 3.1. Introduction

In order to gain trust in a protein structural model it is necessary to test predictions made from the model experimentally. The temptation must be resisted to over-interpret an *in silico* model until some confidence is gained as to how accurately it represents the true structure. The results of experimental investigations can then be used to further refine or modify the model if necessary. Iterating this process then allows for bolder predictions to be made from such a model and thus increase it's usefulness (Flower, 2002).

The human form of AdoMetDC is stimulated by putrescine (Pegg, 1984), whereas the plant enzyme is not (Xiong *et al.*, 1997), and the *Trypanosoma* enzyme is only stimulated by much higher concentrations of putrescine (Clyne *et al.*, 2002). Compared to the human enzyme, the crystal structure of the plant enzyme reveals the presence of certain mutations that may result in stimulation by internal residues: Arg18 and Arg114 occupy the region where a putrescine would bind (Bennett *et al.* 2002, Fig. 3.1). Furthermore, replacing the residue corresponding to Arg18*pot* (equivalent of model Arg11) in *Trypanosoma* AdoMetDC with Leu results in loss of activity (Clyne *et al.*, 2002).

In the previous chapter the homology modelling of malarial AdoMetDC is described. This model was used to determine possible reasons why the malarial enzyme is not stimulated by the polyamine putrescine. Briefly, the model binding site was seen to lack corresponding residues which had been identified as being important for the binding and stimulatory effect of putrescine in humans. However, in the model, a set of charged residues is conserved which has been suggested to transmit the effect of putrescine binding to the active site (Fig. 3.1, Ekstrom *et al.* 2001; Tolbert *et al.* 2001). Furthermore, the model carries positively charged residues in the vicinity where putrescine would bind. Therefore it is suggested that these residues may take over the function of putrescine.

These predictions regarding *P. falciparum* AdoMetDC are interesting for a number of reasons. Firstly, they not only concern the active site, which is highly conserved in terms of residue composition with the human enzyme (Section 2.4.3.2), but also include more diverged regions of the model which are further

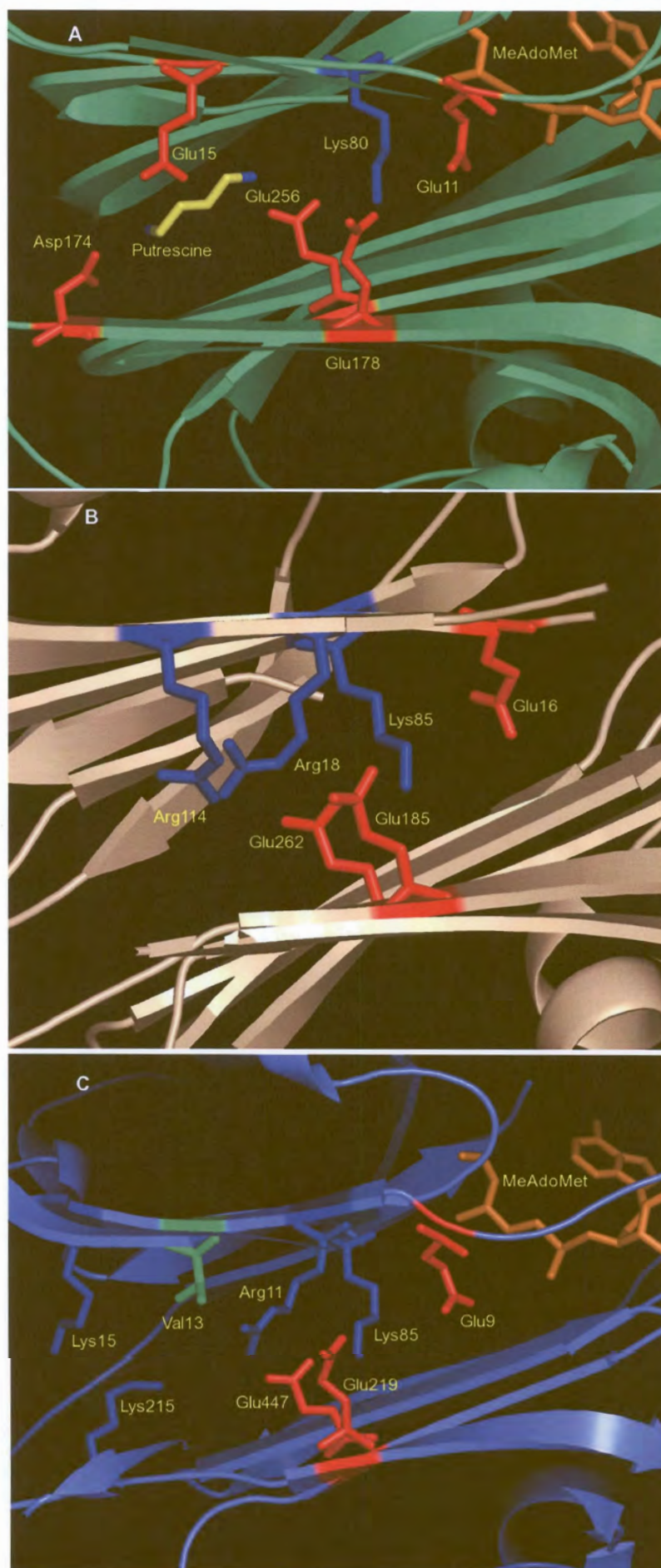


Figure 3.1: Putrescine charge networks. (A) Human (PDB ID: 1I7B). (B) Potato (PDB ID: 1MHM). (C) *P. falciparum* model. Associated acidic residues (red), basic residues (blue) and hydrophobic residues (green) are indicated. MeAdoMet (orange) and putrescine (yellow, blue) are also shown.



removed from the active site (Section 3.4.1). Therefore to target these residues experimentally may help verify the validity of the model's structure in regions further removed from the active site. Of great concern during the modelling process was the low sequence identity between the *P. falciparum* target and human and potato templates ( $\pm 20\%$ ). Importantly, greater confidence may be gained about correctness of the sequence alignment used to build the model. To test whether internal residues could be playing a putrescine-like role in malarial AdoMetDC, the effect of substituting the identified positive residues (Arg11, Lys15, Lys215) with amino acids that should affect enzyme activity was determined. The main objective of this aspect of the study was to determine by means of site-directed mutagenesis on recombinantly expressed enzyme whether internal residues might function as putrescine in *P. falciparum* AdoMetDC.

## 3.2. Methods

### 3.2.1. *In silico* putrescine docking

Reasons for the lack of putrescine stimulation in *P. falciparum* AdoMetDC, seen for both cleavage and activity of the human enzyme, were investigated. For initiating putrescine docking studies the same model subjected to minimisation described was used (Section 2.2.4). The substrate analogue MeAdoMet was added as before. From this a mutated model was generated in order to mimic the human putrescine binding site. The following mutations were introduced into the model using INSIGHTII: Lys215Asp and 11-R-V-K-15  $\rightarrow$  11-L-E-W-15. Rotamer studies in INSIGHTII of the  $\chi$ -sidechain angles demonstrated these new residues were already in a fairly energetically favourable conformation. The mutated model was thus a chimera of the *P. falciparum* and human enzymes. Putrescine was inserted into the putative binding site of each model, via superimposition on the human template bound with putrescine. Hydrogens were added as described (Section 2.2.4). Putrescine was assumed to bind with protonated amines. Energy minimisation was carried out on each of these models, as described until a maximum derivative of 1 of  $kcal.mol^{-1}.\text{\AA}^{-1}$  was reached. The models thus generated were then subjected to docking with the SA-DOCKING module of INSIGHTII. The binding site was defined as all complete residues within 6.5  $\text{\AA}$  of putrescine. Default parameters were used, except that 30 initial putrescine conformations were generated for each model. Binding energies were calculated with the CFF91 forcefield. The chimeric and unmutated models thus generated with the lowest energies were subjected to brief energy minimisation using previous parameters until a maximum derivative of 10  $kcal.mol^{-1}.\text{\AA}^{-1}$ .

### 3.2.2. Construction of putrescine-like mutants

#### 3.2.2.1. Mutagenesis of wild-type bifunctional AdoMetDC/ODC plasmid construct

The wild-type bifunctional gene was previously cloned (Müller *et al.*, 2000; Wrenger *et al.*, 2001) into the pASK-IBA3 expression vector (Institut für Bioanalytik, Göttingen Germany). The plasmid was

extracted by mini-prep from DH5 $\alpha$  *Escherichia coli* cells using the NucleoSpin Plasmid® kit (Macherey Nagel®, Germany). The plasmid's identity was confirmed with restriction enzyme digestion using *Xba*I (10U, New England Biolabs (NEB), USA) and *Hind*III (10U, NEB) at 37° C for 1.5 hr. The digested products were separated by electrophoresis on a 1% agarose gel in TBE buffer (0.09 M Tris, 0.09 M boric acid, 0.02 M EDTA, pH 8.0) at 4-7 V.cm<sup>-1</sup>. Ethidium bromide intercalator (500  $\mu$ g/l) was included in the gel solution for DNA visualisation under UV light.

The extracted plasmid was used as template to construct the putrescine related mutants. Site-directed mutagenesis was carried out according to the protocol of the QuikChange® mutagenesis kit (Stratagene®). The mutants and their primers are listed in table 3.1. The following protocol was followed for PCR: initial incubation of 95° C for 30 sec, was followed by 18 cycles of 95° C for 30 sec, 55° C for 1 min and 68° C for 15 min. Each reaction (50  $\mu$ l) contained 50 ng of plasmid, 125 ng each of forward and reverse primers, 2 mM of each dNTP and 6 U of *Pfu* polymerase (Promega®, USA). After PCR the mixture was incubated with *Dpn*I (40 U, NEB) for 2 hr at 37° C. The product was stored at -20° C until further use.

Table 3.1: AdoMetDC/ODC mutant primers. Mutations are italic.

Mutation	Primers	5' → 3'
Arg11Leu	Forward	GGA-ATT-GAA-AAA- <i>TTA</i> -GTT-GTG-ATC-AAA-TTA-AAG-G
	Reverse	C-CTT-TAA-TTT-GAT-CAC-AAC- <i>TAA</i> -TTT-TTC-AAT-TCC
Lys15Ala	Forward	GG-GTT-GTG-ATC- <i>GCA</i> -TTA-AAG-GAG-AG
	Reverse	CT-CTC-CTT-TAA- <i>TGC</i> -GAT-CAC-AAC-CC
Lys215Ala	Forward	GCT-TCT-ACG-TTT- <i>GCA</i> -TTC-TGT-TCG-G
	Reverse	C-CGA-ACA-GAA- <i>TGC</i> -AAA-CGT-AGA-AGC

For each PCR product (mutant), 10  $\mu$ l was transformed into DH5 $\alpha$  *E. coli* and Top10® competent *E. coli* (Invitrogen®) using heat shock: 100  $\mu$ l cells were incubated on ice for 1 hr, heat shocked at 42° C for 90 sec, then incubated on ice for 2 min. The cells were then transferred to 1 ml of NYZ+ (1% w/v casein hydrolysate, 0.5% w/v yeast extract, 0.5% w/v NaCl, 0.0125 M MgCl<sub>2</sub>, 0.0125 M MgSO<sub>4</sub>, 0.02 M glucose, pH 7.5) medium and grown for 1 hr at 37° C with shaking ( $\pm$ 200 rpm). The cells were centrifuged at 16000g (Eppendorf® 5415 table top centrifuge, Germany) for 1 min. The pellet was resuspended in Luria-Bertani medium (LB: 1% w/v tryptone, 1% w/v NaCl, 5% w/v yeast extract) and plated onto LB-agar plates with 100  $\mu$ g.l<sup>-1</sup> Ampicillin. The plates were grown overnight at 37° C.

For each mutant 5 colonies were picked and inoculated into 4 ml LB medium with 50  $\mu$ g.l<sup>-1</sup> Ampicillin and grown overnight at 37° C with shaking (200 rpm). The cells were pelleted at 3000g for 16 min (Heraeus Megafuge 1.0R®, rotor 2705). For clone preservation 700  $\mu$ l culture was stored at -70° C in 30% glycerol (final concentration). Each plasmid was then extracted from the remaining medium by mini prep (described above) and then subjected to restriction digestion and electrophoresis as described above.

### 3.2.2.2. Sequencing of putrescine-like mutants

Mutations were confirmed using the Big Dye® automated sequencing kit version 2.0. For the Arg11Leu and Lys15Ala mutants the StrepTag® (IBA®) sense primer was used: 5'-AGTGAATGAATA-



GTTCGAC-3'. The Lys215Ala mutant was sequenced with the reverse primer 5'-GGAAGGCTTTCTTT-ATTA-3'. Two isolates were sequenced for each of the Arg11Leu and Lys15Ala mutants. All five isolates of the Lys215Ala mutant were sequenced. For each sequencing reaction  $\pm 1 \mu\text{g}$  of plasmid DNA was incubated with 10 pmol primer and 4  $\mu\text{l}$  reaction mix as per manufacturer instructions. The following cycling parameters were used: 96° C for 10 sec, 48° C for 5 sec, 60° C for 4 min for 26 cycles. To the cycling reaction the following was added for cleanup: 4 volumes HPLC-grade water,  $\frac{1}{2}$  volume 3M Sodium Acetate and 12.5 volumes 100% ethanol were added. The mixture was centrifuged at 16000g for 15 min and the supernatant discarded. The pellet was washed with 12.5 volumes (PCR reaction) freshly prepared 70% v/v ethanol. This was followed by centrifugation for 5 min at 16000g and the supernatant discarded. This step was repeated if required. The pellet was dried at 37° C for 10 min and stored at -20° C. Sequencing was carried out in an ABI 377® Automated Sequencer (ABI, USA).

### 3.2.3. Recombinant expression

For each expression ODC and AdoMetDC deficient EWH331 *E. coli* (kindly provided by Dr H. Tabor, Hafner *et al.* 1979) were transformed with  $\pm 50$  ng plasmid wild-type or mutant plasmid as described in section 3.2.2.1. The Strep-tag II fusion protein system was employed (IBA). One colony was picked and grown to saturation overnight in 10 ml LB medium with 0.05  $\text{g}\cdot\text{ml}^{-1}$  Ampicillin (LB-Amp) at 37° C with shaking (200 rpm). This culture was then inoculated into 1 l LB-Amp medium and grown at 37° C with shaking until  $A_{600}$  reached approximately 0.5 units (logarithmic growth phase). Expression was induced by adding anhydrotetracycline (0.2  $\mu\text{g}\cdot\text{ml}^{-1}$ , Institut für Bioanalytik, Germany). The cultures were shaken overnight at room temperature. The cells were pelleted at 9700 g for 10 min 4° C (Sorvall Superspeed® RC2B, rotor SLA 1500) and the pellet frozen for storage (-20° C). The cells were allowed to thaw on ice and resuspended in 20 ml buffer W (1 M Tris-HCl, 1 mM EDTA, pH 8.0). The cells were incubated on ice for 30 min with 0.1 mg lysozyme. Phenylmethylsulphonyl fluoride (PMSF) protease inhibitor was added to 0.1 mM final concentration. The cells were then sonicated for 10 cycles of 30 sec pulsed sonication followed by  $\geq 1$  min incubation on ice-water (Branson® sonifier 250, settings: output control 2, duty cycle 90). The cell debris was then removed by ultracentrifugation at 100000g for 1 hr at 4° C (Centrikon T-1065®, rotor TFT50.38). The recombinant protein was purified as follows: the supernatant was loaded onto 1 ml bed volume of StrepTactin® Sepharose. The flow-through was reloaded 2-4 times. The column was washed with 15 volumes buffer W, and the protein eluted with 5 volumes buffer E (buffer W + 2.5 mM desthiobiotin) and collected in 1 ml fractions. The column was then regenerated with 15 volumes buffer R (buffer W + 1 mM hydroxyazophenyl benzoic acid) and washed with 10-15 volumes buffer W. Separate columns were used for the wild-type protein and each mutant. Protein concentrations were determined according to the method of Bradford (Bradford, 1976), using calibration curves constructed with bovine serum albumin. Expression was analysed by SDS (Sodium dodecylsulphate) polyacrylamide gel electrophoresis (PAGE), using a 5% stacking gel (5% polyacrylamide, 0.19 M Tris, pH 8.8, 1% SDS, 1% ammonium persulphate, 0.1% TEMED) and a 6% running gel (6%

polyacrylamide, 0.38 M Tris, pH 8.8, 1% SDS, 1% ammonium persulphate, 0.1% TEMED). Protein samples were denatured at 100° C for 5 min with 1 volume denaturing buffer (0.15 M Tris, pH 6.8, 1.2% SDS, 30% glycerol, 15%  $\beta$ -mercaptoethanol, 0.18 mg.l<sup>-1</sup>, bromophenol blue) and separated on the gel at 90V.

### 3.2.4. Enzyme Assays

AdoMetDC activity was monitored by trapping of CO<sub>2</sub> from S-adenosyl-L-[methyl-<sup>14</sup>C]methionine. The assay mixture at pH 7.5 contained in 250  $\mu$ l: 2-20  $\mu$ g protein, 50mM KH<sub>2</sub>PO<sub>4</sub>, 1 mM EDTA, 1 mM DTT, 0.1 mM AdoMet and 12.5-25.0 nCi of labelled AdoMet (58 mCi/mmol, Amersham Pharmacia Biotech). The entire reaction mixture was prepared on ice lacking either protein or substrate mix. The reaction was initiated by addition of the missing component and incubated at 37° C for 15 or 30 min. The released CO<sub>2</sub> was trapped on filter papers (1.5×1.5 cm) coated with 40  $\mu$ g Solvable® (PerkinElmer). The reaction was then stopped by adding 500  $\mu$ l 30% trichloroacetic acid, and the remaining labelled CO<sub>2</sub> was diluted by the addition of 500  $\mu$ l 0.1M NaHCO<sub>3</sub>. The filter papers were counted in 4ml Ultima Gold XR® (Packard) scintillation fluid in a Packard 2000CA Tricarb Liquid Scintillation Analyser. Specific activity is expressed in nmol.mg.min<sup>-1</sup> calculated according to the following equation:

$$\frac{\text{CPM} \times \text{nmol substrate}}{[\text{mg protein}] \times \text{minutes} \times \text{total CPM}}$$

Duplicate assays were performed for each time point (15 and 30 min), to yield four specific activities for a protein extract. If necessary, and the amount of protein extract permitting, entire assay runs were repeated.

## 3.3. Results

### 3.3.1. Putrescine docking

#### 3.3.1.1. Comparison with human and model residues

A number of residues have been implicated in putrescine stimulation in the human enzyme (Table 1.1). Those residues which connect the putrescine binding site with the active site are conserved in the model (Table 3.2). Specifically, Glu11<sub>hum</sub> in the active site is thought to be connected to the putrescine binding site via Lys80<sub>hum</sub>, which in turn is in close proximity to Glu178<sub>hum</sub> and Glu256<sub>hum</sub>. The corresponding model glutamate residues are Glu9, Glu219 and Glu447, respectively. The corresponding model residue for Lys80<sub>hum</sub> is Lys85. A number of other residues that have been experimentally verified or constitute the human putrescine binding site are substituted in the model.



Table 3.2: Residues associated with putrescine stimulation. Substitutions/differences between species are indicated in italics.

Human	Model	Human	Model	Human	Model
Glu11	Glu9	<i>Asp174</i>	<i>Lys215</i>	<i>Thr176</i>	<i>Cys217</i>
Lys80	Lys85	<i>Leu13</i>	<i>Arg11</i>	<i>Phe285</i>	<i>Tyr478</i>
Glu178	Glu219	<i>Asp15</i>	<i>Val13</i>	<i>Ser109</i>	<i>Tyr142</i>
Glu256	Glu447	<i>Trp17</i>	<i>Lys15</i>		

### 3.3.1.2. Putrescine docking

In order to elucidate the lack of putrescine stimulation for the *P. falciparum* enzyme, docking studies were undertaken with putrescine in a binding site defined by superimposition with the human structure. A number of acidic residues are located in the human site, and their requirement for putrescine mediated effects has been borne out by site-directed mutagenesis for the human enzyme, suggesting that bound putrescine is positively charged at the amine groups. The docking was repeated for a chimeric model that more closely resembled the mammalian structure. The docking studies yielded structures of lower energy for the chimeric model than for the unmutated model. There is a marked difference in the value of the intermolecular binding energies however. The lowest energy for the mutated model was  $-853.231 \text{ kcal.mol}^{-1}$ , while that for the unmutated model was  $-498.808 \text{ kcal.mol}^{-1}$ , suggesting a more favourable interaction. The orientation of the lowest energy putrescine structures within the mutated model was similar to that within the human template, with the aliphatic backbone running parallel to the  $\beta$ -strands. The lowest energy structures of the unmutated model were all orientated  $45^\circ$ - $90^\circ$  with  $\beta$ -sheets (Fig. 3.2). The chimeric model also displayed less backbone deviation than the unmutated wild-type models from the templates (Table 2.2).

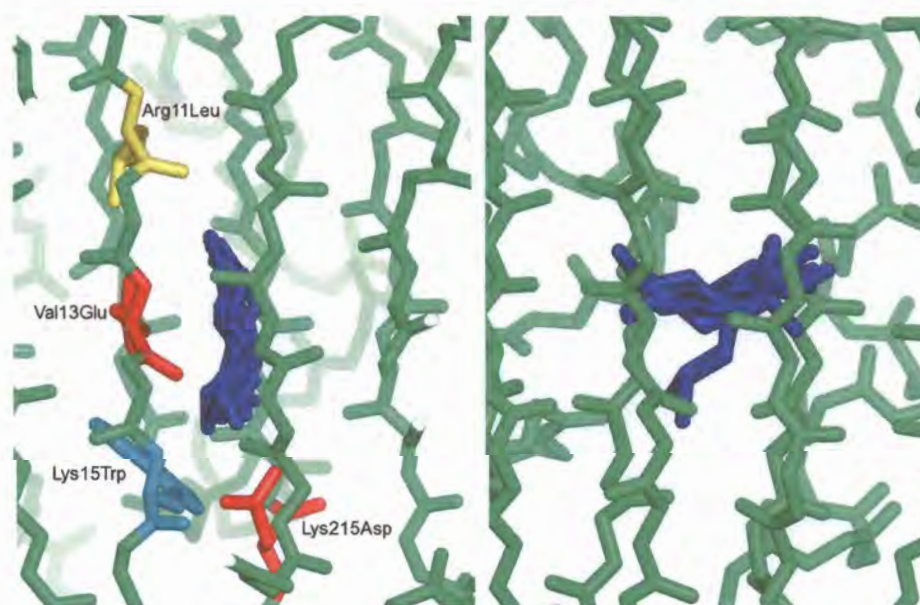


Figure 3.2: Orientation of putrescine. Chimeric (left) and wild-type (right) models. The ten best (lowest energy) conformations of putrescine (blue) are shown for each. *In silico* mutations for the chimeric model are separately coloured.

### 3.3.2. Mutagenesis of wild-type bifunctional AdoMetDC/ODC

The initial extraction of the bifunctional plasmid was successful and restriction digestion resulted in the expected band sizes (Fig. 3.3). The extracted plasmid was thus suitable to be used for site directed mutagenesis. All PCR products of the various mutants were successfully transformed into *E. coli* DH5 $\alpha$  and the plasmids successfully extracted from 5 clones for each mutant (Fig. 3.4). Similar restriction patterns were obtained for most of the extracted plasmids. For each clone sequenced the mutation was positive. One clone for each mutant was subsequently used for expression.

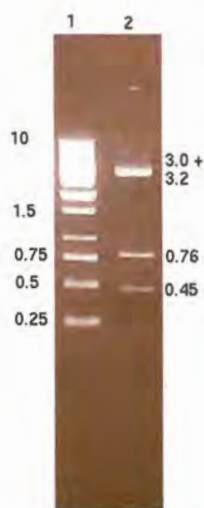


Figure 3.3: *Xba*I and *Hind*III restriction of AdoMetDC/ODC cloned into pASK-IBA3 (right). Lane 1: 1 kb ladder (Fermentas ®). Lane 2: digestion of expression plasmid. Sizes are given in kbp.

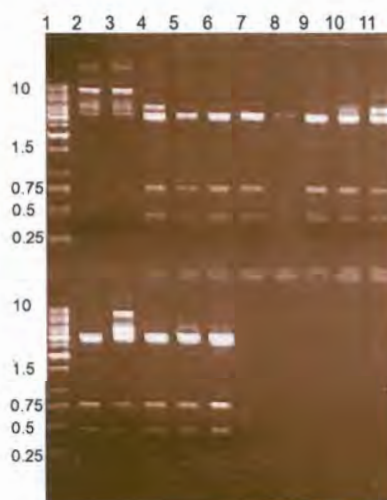


Figure 3.4: *Xba*I and *Hind*III restriction of Arg11Leu (lanes 2-6 top), Lys15Ala (lanes 7-11 top) and Lys215Ala (lanes 2-6 bottom). Marker lanes = lane 1 top and bottom. Incomplete digestion occurred in the first two lanes of Arg11Leu. Sizes are given in kbp.



### 3.3.3. Expression of mutant AdoMetDC/ODC

Overall, the Arg11Leu mutant resulted in an almost inactive protein. The Lys15Ala and Lys215Ala mutants tended to be less active compared to the wild-type. In order to statistically analyse the results, individual activities were calculated for selected experiments. The results indicate that Lys15Ala and Lys215Ala each inactivate the protein by  $\pm 50\%$  (Table 3.3, Fig. 3.5).

Table 3.3: Average relative activities (% wild-type control). Each value is the average from the four individual activities (duplicate assays for 15 and 30 min). The assay run was repeated for experiment 2 (A & B). nd: not determined.

Experiment	Relative activities (%)		
	Arg11Leu	Lys15Ala	Lys215Ala
1	10.14	65.82	61.49
2A	8.05	48.51	34.85
2B	5.44	58.33	44.8
3	2.2	36.9	15.41
4	nd	46.12	90.86

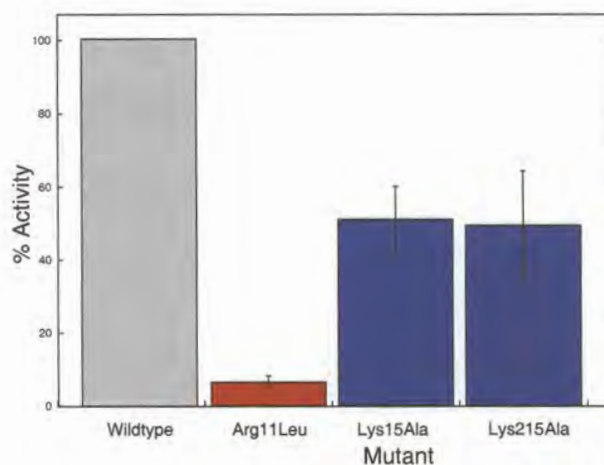


Figure 3.5: Effect of mutations on AdoMetDC activity. Mean relative activities and standard deviations for each mutant are derived from all observations used to construct Table 3.3. Relative activity was calculated from specific activity for each mutant relative to the wild-type activity from that particular assay run. The wild-type activities were normalised to 100%.

## 3.4. Discussion

### 3.4.1. Putrescine docking

The processing and activity of certain AdoMetDC enzymes is stimulated by the presence of putrescine, but not so for *P. falciparum* (Wrenger *et al.*, 2001). Mutation studies have demonstrated that Glu11, Glu15, Asp174, Glu178 and Glu256 are necessary in the human enzyme for stimulation of processing and/or activity by putrescine (Stanley and Pegg, 1991; Stanley *et al.*, 1994; Ekstrom *et al.*, 2001). One putrescine molecule binds each monomer of the human crystal structure. The putrescine binding site is 15-20 Å removed from the active site, and it has been suggested that putrescine exerts its effect via an

internal network of charged residues (Ekstrom *et al.*, 2001). Specifically, Glu11*hum* in the active site is thought to be connected to the putrescine binding site by Lys80*hum*, Glu178*hum* and Glu256*hum*. This series of residues is observed to form an alternating sequence of negative and positive charges. The corresponding model residues are Glu9, Lys85, Glu219 and Glu447, respectively. Lys80*hum* is conserved across all species except for *Leishmania donovani*, *Trypanosoma cruzi* and *T. brucei* where it is replaced by an isoleucine (Fig. A.1, App. A). Glu178*hum* is represented by Glu218 in the model, and appears to be replaced by a serine in *Trypanosoma* and *Leishmania*. *T. cruzi* exhibits stimulation of enzyme activity by putrescine, albeit at significantly higher concentrations, and by an apparently different mechanism of increasing  $V_{max}$  instead of lowering  $K_m$  (Kinch *et al.*, 1999). Therefore it appears that for these parasites the lack of putrescine stimulation may be due to the partial loss of residues needed to transmit the charge effect. Asp174*hum* on the other hand is less conserved in other species, appearing as hydrophobic residues in plant sequences and as either Lys or Asn in *Plasmodium sp.* In the case of the model this is Lys215. Thus it appears that *P. falciparum* could have a similar transduction mechanism for putrescine-like effects, despite a lack of observable putrescine stimulation (Wrenger *et al.*, 2001).

In the human enzyme the binding of putrescine itself is mediated partly by interactions between the positive amine terminals with Glu15*hum*, Asp174*hum*, Glu178*hum* and Glu256*hum*. The corresponding model residues are Val13, Lys215, Glu274 and Glu447 (Fig. 3.1). Therefore, two negatively charged residues that could potentially interact with putrescine are absent, one of them being replaced by a positive Lys residue. Further inspection of where putrescine would bind revealed the presence of two further positive residues (Arg11 and Lys15) that would approximately occupy the region of the putrescine amine terminals. Therefore at one terminal there is an Arg residue, and at the other terminal there are two Lys residues. Any interaction with putrescine is therefore expected to be unfavourable due to repulsion between the amine ends of putrescine and these particular residues.

In order to elucidate the lack of putrescine stimulation for the *P. falciparum* enzyme, docking studies were undertaken with putrescine in a binding site defined by superimposition with the human structure. This was repeated for an *in silico* mutated malarial model which more closely resembled the mammalian structure. There was a marked difference in the value of the intermolecular binding energy between putrescine and the protein ( $\pm 400 \text{ kcal.mol}^{-1}$ ). A more energetically favourable interaction was obtained for the mutated model. The orientation of the lowest energy putrescine structures within the mutated model was also similar to that within the human template, with the aliphatic backbone running parallel to the  $\beta$ -strands. The lowest energy structures of the unmutated model were all orientated  $45^\circ$ - $90^\circ$  with  $\beta$ -strands (Fig. 3.2). These results suggest that in order for putrescine to stimulate the *P. falciparum* enzyme, it would have to be engineered to enable favourable putrescine binding. In the mutated model the 11-R-V-K-15 motif of a  $\beta$ -strand is replaced with 11-L-E-W-15 which is seen in mammalian sequences, where the middle Glu residue corresponds with Glu15*hum*. The side chains for these residues are orientated towards the cavity occupied by putrescine. Notably Arg11 and Lys15 are found in the vicinity of the putrescine amine groups. These side chains are assumed to be positively charged, as are



the putrescine amino groups. It is thought that these residues may simulate the ends of a putrescine molecule, while the intervening Val residue may simulate the putrescine backbone. In the potato structure a similar situation is observed where Arg18*pot* occupies the same position as Arg11. Arg18*pot* has been suggested to simulate putrescine together with Arg114*pot* and other residues (Bennett *et al.*, 2002). In the *P. falciparum* model, however, Arg114*pot* is replaced by Phe144. Furthermore, mutating Arg18 of the *T. cruzi* enzyme abolishes putrescine stimulation in that organism (Clyne *et al.*, 2002). The last mutation introduced into the chimeric model replaces Lys215 for an Asp near the end of docked putrescine. The mutant model also displayed less backbone deviation from the templates than the unmutated model (Table 2.2). A lower backbone deviation was also seen between the mutant model docked with putrescine and the original undocked model, than between the wild-type structures with and without putrescine. Thus assuming the wild-type undocked model is closest to the enzyme native structure, it would seem that the mutant model adopts the more correct structure of the two docked models, and may well be functional. Overall, these results suggest that internal residues may play the role of putrescine seen in other species. Furthermore, the *in silico* mutations described, if carried out in practice, may allow binding of putrescine and convert the malarial enzyme to a putrescine stimulated-enzyme.

### 3.4.2. Mutagenesis

The recombinant plasmid with bifunctional AdoMetDC/ODC was successfully isolated and mutated to incorporate the Arg11Leu, Lys15Ala and Lys215Ala point mutations in separate constructs. Arg11 was mutated to Leu, since Leu is the corresponding residue in humans, and similar experiments on the *Trypanosoma* AdoMetDC demonstrated that the corresponding residue was required for activity (Clyne *et al.*, 2002). The Arg11Leu mutation essentially inactivated the protein. Relative activity for the Arg11Leu mutants never exceeded 12%, and the activity was hardly ever more than the negative control. Thus it is concluded that Arg11 is required for correct functioning of the AdoMetDC domain of the bifunctional enzyme. According to the model, Arg11 occupies a similar position to one amine terminal of putrescine seen when superimposed with the human crystal structure of AdoMetDC. Both Arg11 and putrescine are in close contact with Glu residues in their respective structures, and are expected to be positively charged. As described in section 3.4.1 an identical network of charged residues connects the active sites with either putrescine or Arg11. A similar situation is seen in the potato crystal structure, and mutation of the corresponding residue in *Trypanosoma* AdoMetDC identified by alignment inactivates that enzyme. Based on this combined evidence it is argued that Arg11 simulates the role of putrescine in *P. falciparum* AdoMetDC to give a constitutively active enzyme, much as was found in the potato AdoMetDC (Bennett *et al.*, 2002). No effect on ODC activity was expected, since the targeted residues are all part of the AdoMetDC core according to the model. The effect of these mutations on ODC was therefore not determined, although this is still open to future experimentation.

Whereas Arg11 occupies one amine terminal of putrescine in the model, Lys15 and Lys215 are predicted to occupy the other amine terminal. Since both Lys residues are expected to be positively charged,



either or both may potentially simulate putrescine. The results of site directed mutagenesis indicate loss of each residue results in  $\pm 50\%$  less activity. Initially Lys15 was expected to be the residue more likely to affect enzyme activity, since it occurs on the same  $\beta$ -strand as Arg11 and is part of a region of greater sequence identity with the templates. Lys215 occurs in a more divergent region, and it was anticipated that its position in the model may be the result of misalignment in model construction. Therefore, Lys215 was expected to be less likely to affect AdoMetDC activity. Both residues were mutated to Ala, since this is physicochemically a substantially different residue, and no studies targeting cognate residues in other organisms were found in the literature. For both residues,  $\pm 50\%$  inactivation of the enzyme was observed, however, the results for Lys15Ala were more reproducible. A couple of reasons are suggested for this. Firstly, the total protein concentration varied significantly for different extracts, although similar amounts of total protein were included in the assays. As a general rule, greater activity was observed with extracts of higher protein concentrations. It is well documented that proteins can alter activity due simply to the presence of other macromolecules causing macromolecular crowding (van den Berg *et al.*, 1999; Minton, 2001). Furthermore, there may be specific stabilising interactions between the extracted proteins. Secondly, no dialysis was performed on the protein extract in order to convert the protein environment to that of the assay buffer. Since different volumes were added to accommodate the different protein concentrations, slightly different amounts of expression buffer salts were present in the assay reactions, which may have affected the enzyme activity. It nonetheless appears that both Lys15 and Lys215 are required for the optimal functioning of *P. falciparum* AdoMetDC. Furthermore, since replacement of each mutant results in approximately 50% less activity, it is suggested that these residues may function together to simulate one amine terminal of putrescine in *P. falciparum* AdoMetDC. These two residues are also not connected by the same charge network that joins Arg11 to the active site. Instead Lys15 is separated from Arg11 by a Val residue in the same residue. It was already suggested that this R-V-K may simulate the presence of putrescine. It is therefore further suggested that the effects of these residues are at least partially mediated by the intervening Val13 residue. In summary, the enzyme appears to have mechanisms to simulate the functioning of putrescine, that are mediated by internal residues. Because the targeted residues were correctly identified to affect activity from the model, these results also partially validate the model and the sequence alignment used to construct it. Further experiments to be carried out would be the construction of the double Lysine mutant, and the construction of the putrescine-binding human-parasite chimera modelled *in silico* (Section 3.4.1), to see if putrescine binding can be engineered into this protein.

The fourth chapter describes the use of the model to identify potential novel inhibitors for *P. falciparum* AdoMetDC *in silico*, and the experimental testing of some of these.



## Chapter 4

# Model guided inhibitor screening of malarial

## AdoMetDC

### 4.1. Introduction

#### 4.1.1. *In silico* ligand docking

Intuition suggests that given the 3D structure of a protein it should be possible to discover a novel ligand without resorting to biochemical experiments. This is the premise on which ligand docking is based (Shoichet *et al.*, 2002). Fitting a potential ligand into a protein binding site computationally is referred to as the “docking problem”, and has much in common with the so-called “folding problem” of *ab initio* protein structure prediction (Halperin *et al.*, 2002). The reason for describing this as a “problem” is not trivial. While the physics of chemical interactions have been well understood for about 70 years, the simulation of these interactions *in silico* constitutes a very computationally expensive process. As an example the number of possible conformations for a compound with 10 rotatable bonds and 3 minima per bond yields 59049 conformations. Increasing this to six minima per bond results in over  $3.48 \times 10^9$  conformations (Halperin *et al.*, 2002). At the time of writing, a typical desktop CPU can perform  $2 \times 10^9$  operations per second, many thousands of which would be required to evaluate one molecular conformation. Thus, an exhaustive fine grained search is not always computationally feasible, particularly if a large library of tens of thousands of compounds is to be screened. A number of docking methods exist and will be briefly outlined here.

The first aspect of the docking problem is evaluating the interaction between ligand and protein. This is referred as the scoring problem. Over the last 30 years a number of methods have been invented to tackle this problem (Halperin *et al.*, 2002; Taylor *et al.*, 2002; Shoichet *et al.*, 2002). The most accurate of these are simply extensions of force field based methods used for performing minimisation and dynamics of chemical systems *in silico*. Whereas forcefield based scoring schemes are more accurate, they are computationally expensive, and therefore only suitable if a small number of ligands is under investigation. If a database of many thousands or millions of compounds has to be evaluated however, a faster scoring scheme is required. For this purpose knowledge-based and empirical scoring functions have been developed. The former are derived from observed atom-atom contacts observed in protein-ligand

structures, whereas the latter are derived from fits to experimentally determined binding energies. Knowledge based and empirical scoring schemes represent the ligand and/or active site cavity as point grids. These scoring schemes employ elements such as geometric complementarity, contact and overlap checks, counts of hydrogen bonds, counts of un-neutralised charges, total buried surface areas, etc, which are faster to compute.

Forcefield based scoring methods are known to be prone to over-estimating interaction energies. Knowledge-based and empirical functions are less likely to calculate high energies, however, they can suffer from incorrect interpretation of the data from which they were derived. Furthermore, faster scoring schemes are generally combined with less exhaustive searches and thus relevant binding modes are more likely to be overlooked. Recently there has been much motivation for the use of fast scoring functions for initial screening of large databases, followed by finer grained searches on the high scoring hits from the initial search. Another problem of scoring schemes is that the ranking of high scoring compounds is frequently incorrect. Furthermore, for a particular compound the correct binding mode may be identified with a high score but is still not predicted as the most likely. To overcome this, consensus methods that employ multiple scoring schemes are gaining popularity (Krumrine *et al.*, 2003).

Exhaustive searches of ligand binding modes are generally not feasible for reasons discussed. A number of methods have therefore been derived to overcome this problem (Halperin *et al.*, 2002; Taylor *et al.*, 2002; Shoichet *et al.*, 2002). Forcefield based docking methods generally make use of typical minimisation and molecular dynamics procedures. As described above, the degrees of freedom for a moderately sized system are too numerous to allow for exhaustive searching for a local minimum. Minimisation of an energy function is not a simulation of the behaviour of a chemical system, however, it is simply an optimisation that attempts to find a local minimum. Molecular dynamics combines a forcefield with Newton's equations of motion in order to attempt to predict how a chemical system actually moves. Molecular dynamics also obviously lends itself to docking. Minimisation and molecular dynamics are very computationally expensive, however, and therefore are usually only feasible for small numbers of ligands. Other methods to find a minimum include Monte Carlo simulations, or genetic algorithms (e.g. GOLD, AUTODOCK). Monte Carlo simulations refer to methods that explore search spaces by random sampling in order to obtain a representative population of the search space, and thus determine areas of low and high energy. Genetic algorithms make use the concepts of Darwinian selection to optimise the fitness of individual scores associated with molecular conformations.

Searching can be defined into three classes based on the treatment of flexibility. Most methods treat the ligand as flexible or rigid with a rigid binding site. In the rigid case the ligand has only six degrees of freedom (three rotation and three translation) which allows for fast searching. Flexible ligand employs various methods to explore internal degrees of freedom as well. The protein target is usually treated as rigid due to computational constraints, however, recently more methods attempt to add flexibility to the protein as well. This can be done implicitly in the scoring function to allow for so-called soft potentials, i.e. the ligand and protein surfaces are allowed some degree of interpenetration. Protein flexibility can



also be accommodated through the use of rotamer libraries for protein side-chains that generate the most probable conformations, or through procedures typical of minimisation and molecular dynamics. The subject of docking is further reviewed in Halperin *et al.* (2002); Taylor *et al.* (2002); Shoichet *et al.* (2002); Krumrine *et al.* (2003).

This chapter describes the *in silico* docking of compounds from the ACD (Available Chemicals Directory) and the NCI (National Cancer Institute) database in the active site of the *P. falciparum* AdoMetDC model. Some of these compounds were selected for preliminary biochemical screening in order to determine their effectiveness as inhibitors. Inhibitors identified in this manner may serve as lead compounds for new drugs, furthermore, the success of biochemical screening can be used to further gauge the accuracy and potential usefulness of the model.

## 4.2. Methods

### 4.2.1. *In silico* inhibitor screening

*In silico* screening of potential drugs was performed using the LUDI module of INSIGHTII against the National Cancer Institute (NCI), Available Chemicals Directory (ACD) databases of small molecules as well as the database internal to LUDI (BIOSYM). The runs were conducted using a radius of 11 Å, with the position of N3 of the adenine ring of MeAdoMet as the centre of the searches. Due to the importance of the packing of the cyclic planar systems of the substrate between Phe7 and Phe223 for binding in the human enzyme (Tolbert *et al.*, 2001), a scoring function was used since that also includes aromatic-aromatic interactions (*energy\_estimate\_3*) in addition to ionic, hydrophobic and hydrogen-bonding interactions. This was repeated for the malarial model and the human enzyme. Various properties for the high scoring NCI hits, such as the *n*-octanol partition coefficient (*logP*) and potential for inhibition of various enzymes, was predicted on the NCI database website using the PASS functionality (Prediction of Activity Spectra for Substances, Poroikov *et al.* 2003)

### 4.2.2. Test compound solutions

The top 11 distinct compounds were selected from the results of virtual screening of the AdoMetDC model against the NCI database (Table B.3, App. B). The compounds were obtained from the NCI. Stock solutions of the test compounds were made with dimethyl sulphoxide (DMSO) and dimethyl formamide (DMF). The appropriate solvent was first determined by dissolving a single grain and observing on a glass microscope slide. The effect on solubility of dissolving the test compounds in assay buffer and ethanol was also determined using an ordinary light microscope and a Normaski interference microscope. All compounds ( $\pm 5$  mg each) were dissolved in 1 ml DMSO or DMF to make stock solutions, except for compounds 10 and 11 which were dissolved in 1 ml DMF (Table 4.1).

### 4.2.3. Assays

Wild-type bifunctional enzyme was expressed as described above (Section 3.2.3). AdoMetDC assays were performed essentially as described (Section 3.2.4), with reactions started by addition of enzyme extract. The potential-inhibitor stock solutions were diluted 1/10 in assay buffer or ethanol then added to the assay. The final concentrations of the potential inhibitors ranged from approximately equimolar with the enzyme substrate to approximately 100× the substrate concentration. Solvent (DMSO and DMF) controls were also included. Reactions were set up in duplicate for each test compound. All incubation times were for 30 min.

## 4.3. Results

### 4.3.1. *In silico* inhibitor screening

Both the malarial model and the human enzyme were screened against a number of small database molecules. In each case the top 20 scoring molecules were different. The scores were relatively high, with correspondingly low predicted  $K_i$  values (LUDI, using the formula:  $K_i = \exp(-score/100)$ ). The predicted inhibitors are dominated by planar aromatic systems with high calculated  $logP$  values (PASS predictions). Most of the identified compounds were heterocyclic. The top hits identified for the NCI and ACD generally differed, and those identified from screening against the NCI database contained more heteroatoms than for the ACD. The predicted inhibitors identified for the *P. falciparum* enzyme were usually different compared to the corresponding set for the human enzyme. None of the predicted inhibitors resemble the substrate or known inhibitors of human AdoMetDC (Fig. B.1-B.3, App. B). All of the compounds lack the ribose moiety that is seen in the substrate, and only one compound carries a positive charge (Compound 4, Table 4.1). Compounds 5 and 8 do show some resemblance to the adenine portion of MeAdoMet (Table 4.1). Despite the overall lack of resemblance, some of these compounds were predicted to be AdoMetDC inhibitors by the NCI database website using the PASS prediction function. Therefore, enough interest in these compounds remained to test them biochemically. The compounds selected for biochemical assays are shown in Table 4.1. The orientations of the top 6 scoring compounds are shown in Fig. 4.1.

### 4.3.2. Solubility of potential inhibitors

All of the potential AdoMetDC inhibitors dissolved in DMSO except for compounds 10 and 11, which were dissolved in DMF. Compound 6 was still partly in suspension after dissolving in DMSO. Compounds 1, 2, 3, 6, 8, 10 and 11 were observed to recrystallise when diluting 10-fold in assay buffer. Diluting 10-fold in ethanol was more successful, since only compound 6 recrystallised. However, when diluting the ethanol solutions a further 10-fold in assay buffer, compounds 1, 3 and 11 also recrystallised.



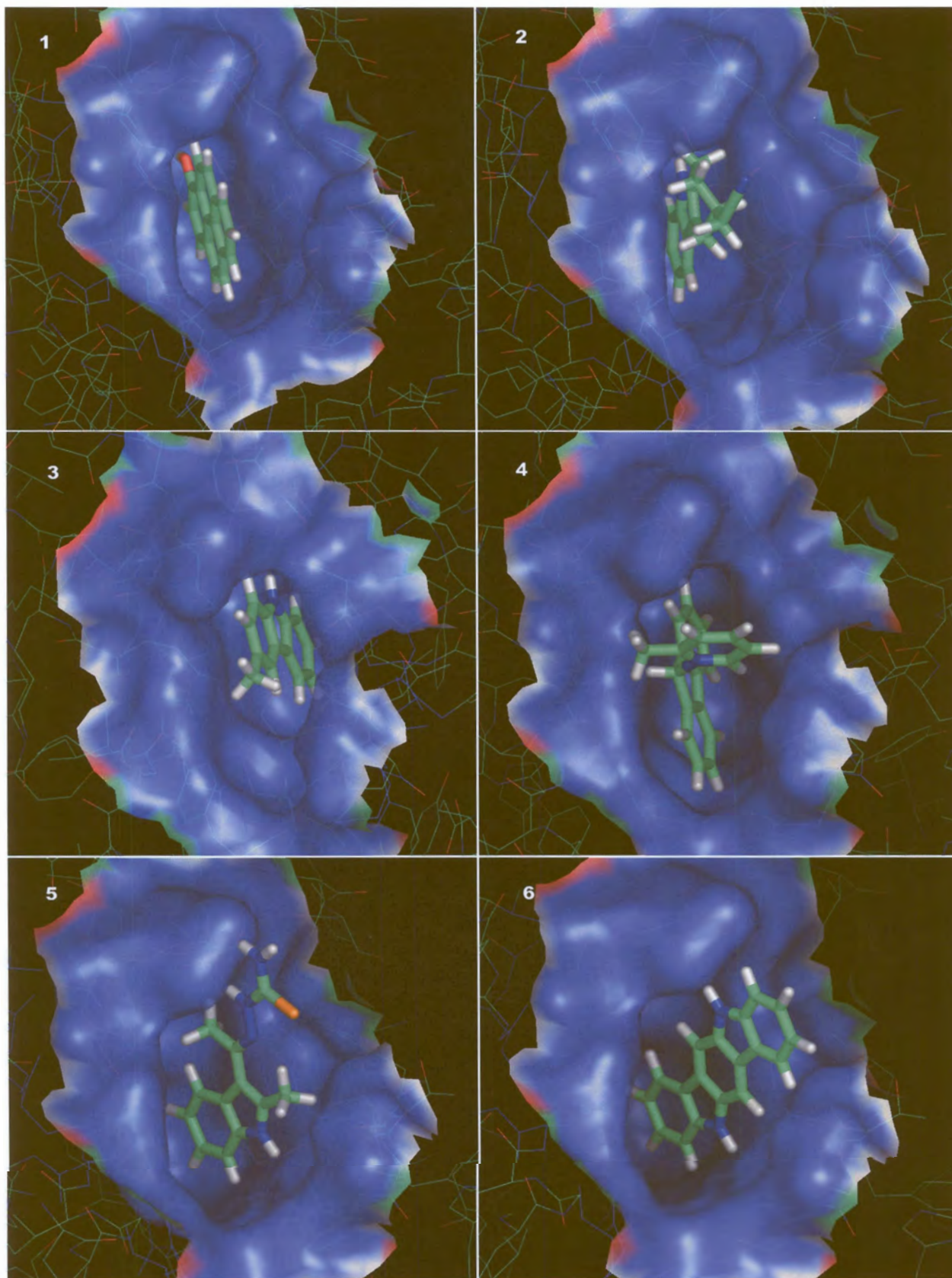

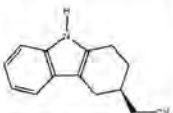
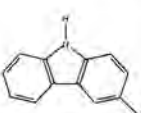
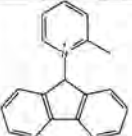
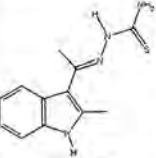
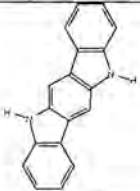
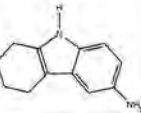
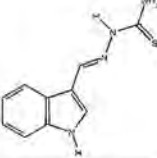
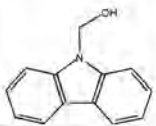
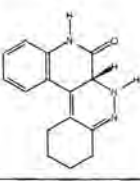
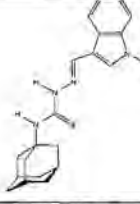


Figure 4.1: Orientations of the top 6 potential NCI inhibitors.

Table 4.1: Potential NCI inhibitors identified for AdoMetDC selected for testing. The mass used to make stock solutions is indicated.

Hit #	1	2	3	4
Structure				
MW ( $g.mol^{-1}$ )	218.25	210.28	181.24	258.34
Predicted $K_i$ (M)	$2.8 \times 10^{-08}$	$7.4 \times 10^{-8}$	$8.1 \times 10^{-8}$	$1.5 \times 10^{-7}$
Hit #	5	6	7	8
Structure				
MW ( $g.mol^{-1}$ )	246.33	256.31	186.26	218.28
Predicted $K_i$ (M)	$1.7 \times 10^{-7}$	$2.0 \times 10^{-7}$	$2.1 \times 10^{-7}$	$2.1 \times 10^{-7}$
Hit #	9	10	11	
Structure				
MW ( $g.mol^{-1}$ )	197.24	253.30	352.50	
Predicted $K_i$ (M)	$2.1 \times 10^{-7}$	$2.2 \times 10^{-7}$	$2.5 \times 10^{-7}$	

#### 4.3.3. Inhibition of AdoMetDC

First, the effect of approximately equimolar concentrations of each potential inhibitor was determined using the standard assay. Each reaction contained either 10  $\mu$ l of test compound stock solution dissolved 10-fold in assay buffer, or 10  $\mu$ l of DMSO or DMF as controls. No marked inhibition was observed. In contrast, most reactions with the potential inhibitor included displayed slightly higher activity (Fig. 4.2 A).

Following this, the effect of approximately equimolar concentrations of each potential inhibitor to substrate was determined at 1/10 the concentrations used in the previous experiment (only radioactively labelled substrate was included). Each reaction contained either 1  $\mu$ l of test compound stock solution dissolved 10-fold in assay buffer or ethanol. The choice of assay buffer or ethanol was based on the solubility results. 1  $\mu$ l each of absolute DMSO, DMF and ethanol were included as controls. Those compounds which had proved insoluble were excluded. Marked inhibition was only observed for compound 10 ( $\pm 50\%$ , Fig 4.2 B).

The next experiment was essentially as in Figure 4.2 B. The effect of approximately equimolar concentrations of each potential inhibitor to substrate was determined at 1/10 substrate concentration. Each reaction contained either 1  $\mu$ l of potential inhibitor dissolved 10-fold in assay buffer or ethanol. Potential inhibitor concentration was approximately equal to substrate concentration. However, the solvent controls were designed to have similar solvent concentrations to the test compound assays. In the previous experiments excess solvent controls had been used ( $\pm 10\times$  compared to test compound assays)



in order to establish for certain that DMSO and DMF were having no effect. For this experiment, 1  $\mu$ l of 10-fold dilutions of DMF and DMSO in either assay buffer or ethanol were included as controls, as well as absolute ethanol. Those compounds which had proved too insoluble were excluded. No marked inhibition was observed (Fig 4.2 C).

Finally, the effect of approximately 10 $\times$  test compound to substrate concentrations of each potential inhibitor was determined at 1/10 substrate concentration. Each reaction contained either 10  $\mu$ l of potential inhibitor diluted 10-fold in assay buffer or ethanol. 10  $\mu$ l of 10-fold dilutions of DMF and DMSO in either assay buffer or ethanol were included as controls, as well as absolute ethanol (equimolar solvent concentration in both solvent controls and test compound assays). Those compounds which had proved too insoluble were excluded. A small degree of inhibition was only observed for compounds 4, 5, 7, 8, 9 and 10 ( $\pm$ 50%, Fig 4.2 D).

In summary the potential inhibitors had little effect on the AdoMetDC activity. Compound 10 was an exception to this and on two occasions displayed noticeable inhibition. More inhibition was observed when a test compound to substrate concentration ratio of  $\pm$ 10:1 was used. The solvent controls appeared to have little effect on AdoMetDC activity.

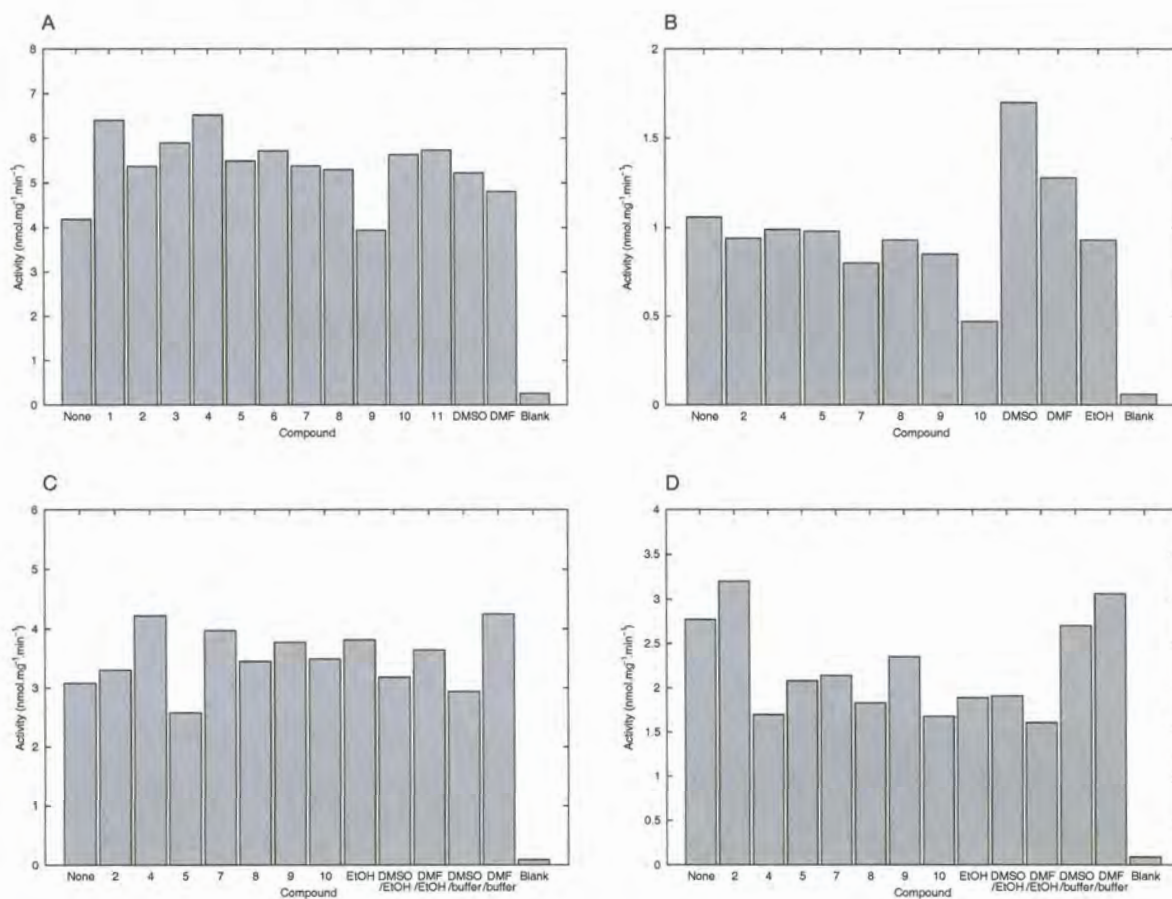


Figure 4.2: Overview of effect of identified inhibitors and solvent controls on AdoMetDC activity. (A) Equimolar test compound and substrate, excess solvent controls. (B) Equimolar test compound and substrate, excess solvent controls, 1/10 substrate/test-compound concentration. (C) Equimolar test compound and substrate, solvent in controls equimolar with test compound assays. 1/10 substrate/test-compound concentration. (D) 10 $\times$  test-compound to substrate concentration, solvent in controls equimolar with test compound assays, 1/10 substrate concentration .

#### 4.4. Discussion

Both the malarial model and the human enzyme were screened against the LUDI (BIOSYM), NCI and ACD databases of small molecules. In each case, the top 20 scoring molecules were different. The scores were relatively high, which correspond with nanomolar predicted  $K_i$  values. This was taken to imply that the host and model are sufficiently different that rational inhibitor discovery using this approach is a worthwhile pursuit. The predicted inhibitors are dominated by planar aromatic systems with high calculated  $\log P$  values. This degree of hydrophobicity is expected to make testing of these compounds difficult in practice, even though the docking results indicate that valid interactions are possible between the predicted binders and the model. None of the predicted inhibitors resemble the substrate or known inhibitors of human AdoMetDC. Most notably no nucleoside-like compounds with a pentose moiety were discovered. A small library of known inhibitors (Fig. 4.1) was also constructed and docked with the model and human crystal structure. In neither case however, was it possible for LUDI to dock any of these compounds. This is thought to be due mostly to incomplete sampling of the chemical space available. Due to software constraints it was only possible to do rigid docking of the molecules for the large databases. The LUDI system is built with fast mass screening of molecules in mind, therefore it is considered necessary to use an algorithm that can conduct a more refined search in order to identify improved hits.

Biochemical assays of selected hits from screening against the NCI database were carried out in order to determine whether any of these compounds warranted further analysis. No reproducible inhibition was observed for assays containing approximately equimolar amounts of substrate and compound. When the ratio of compound to substrate was increased to 10 $\times$ , slightly more inhibition was observed for some compounds. However, a similar degree of inhibition was observed for the ethanol controls, therefore it cannot be concluded that the compounds were responsible for any inhibition. Furthermore, if the lower activity was due to inhibition by the test compounds, this was only at a very high concentration. Since there was no remarkable inhibition at equimolar tests-compound and substrate concentration no follow up kinetic studies were performed for any of the test compounds. Most of the compounds in DMSO stock solutions were seen to precipitate when diluting directly into assay buffer. This was apparently partially alleviated by first diluting in ethanol, then in assay buffer. However, the final high dilution of the compound stock solutions made it difficult to determine under a microscope that no recrystallisation was occurring. Because many of the compounds appeared to precipitate, it is concluded that in many cases no inhibition could have occurred: the compounds would have to be in solution to be able to reach the AdoMetDC active site. The relative insolubilities are in concord with the high predicted  $\log P$  values of the compounds. Lipinski *et al.* (1997) determined a number of properties (Lipinski's rule of 5) that distinguish orally bioavailable drugs from other compounds (excluding compounds that are transported by membrane proteins). Namely, orally bioavailable drugs tend to have  $\leq 5$  hydrogen-bond donors,  $\leq 10$  hydrogen-bond acceptors,  $\log P < 5$ , and MW  $\leq 500$ . These properties to some degree



represent solubility, since insoluble drugs will not be orally bioavailable. However, the rule of 5 is also indicative of a compound's ability to cross the lipid bilayer, a process required for passive absorption of drugs. Therefore, even though compounds may obey the rule of 5, that does not guarantee their aqueous solubility. Of the test compounds, only compound 11 disobeys the Lipinski's rule of 5, with a predicted  $\log P > 5$ . Therefore, most of the test compounds could be drug-like based on these criteria, however, not necessarily soluble. Recently rational drug-design has shifted focus from identification of "drug-likeness" to identification of "lead-likeness". Good lead-like compounds are those considered likely to serve as good scaffolds upon which improvements can be made for creation of new drugs. Lead-like compounds tend to be more hydrophilic and smaller than drug-like compounds (Rishton, 2003). Therefore, screens such as the rule of 5 and those for lead-likeness should be applied in future studies to select compounds that are more experimentally tractable for inhibition studies.

Docking algorithms are imperfect and may miss valid binding modes or incorrectly assign relevance to a particular mode (Halperin *et al.*, 2002; Taylor *et al.*, 2002). This is reflected in the literature whereby compounds predicted to bind *in silico* are nonetheless tested at high (micromolar) concentrations. Therefore the lack of correspondence between the experimental results and the predicted binding does not necessarily invalidate the model. In future, it is suggested that multiple docking algorithms should be applied and a consensus set of compounds chosen for biochemical investigation. Such a set of compounds should also be subjected to more refined searches to provide extra validation.

## Chapter 5

# Concluding Discussion

*P. falciparum* ODC/AdoMetDC represents a unique opportunity to gain further understanding of an aspect of malarial metabolism that is very parasite specific. Both of these activities are key regulatory points in the malarial metabolism of polyamines, and this bifunctional arrangement is so far unique to *Plasmodium sp.* Thus gaining further understanding of this protein may lead to the discovery of novel anti-malarials. At the beginning of this study there was little known about the structure of this domain. The aim of this study was to obtain knowledge of the structure of the *P. falciparum* AdoMetDC domain from the bifunctional enzyme, primarily by using *in silico* methods. It was expected that this would lead to further understanding of the unique structural features of this enzyme, and that this knowledge could be employed to discover potential new inhibitors.

A 3-dimensional model of the AdoMetDC domain was constructed using homology modelling. Using the known crystal structures of the human and potato enzymes, together with a sequence alignment comprising all these proteins a model for malarial AdoMetDC was constructed. This model was subjected to minimisation procedures to relieve unlikely conformations. From the model a number of unique characteristics for the malarial AdoMetDC could be proposed. A summary of key differences between the model and the human enzyme is given in Table 5.1.

Firstly, in the process of constructing this model it was discovered that the bifunctional enzyme is also present in other species of *Plasmodium*. By using the data of the *Plasmodium* genome sequencing project (The *Plasmodium* Genome Database Collaborative, 2001) it was possible to find complete sequences of the bifunctional enzyme for *P. yoelii* and *P. berghei*. From studying the *Plasmodium* sequences a number of predictions about the structure of the AdoMetDC domain could be made. From the sequence alignments it became evident that the AdoMetDC domain contains 3 *Plasmodium*-specific inserts. Secondary structure predictions and structural modelling suggested that the shorter inserts fold into conserved structures and are more important to the functioning of the AdoMetDC domain than the longer insert. The longer insert displayed considerably more variation between the *Plasmodium* sequences, and as a result of this it was suggested that this insert could be partially dispensed with. Deletion mutagenesis of a region containing this insert contradicted this prediction (Birkholtz *et al.*, 2004). However, this deletion was performed before any structural model of AdoMetDC was available. The modelling conducted in this study indicates that this deletion also contained core folding regions, and thus in hindsight is expected



Table 5.1: Summary of main differences between the malarial (model) and host enzymes.

Model	Human
Substituted active site residues	
Gly3	His5
Thr416	Asn224
Tyr435	Ile244
Ser436	Thr245
Active site cavity	
Large cavity near pyruvoyl residue	Absent
	Cavity near methyl group of substrate appears larger
Putrescine stimulation	
In the model, positively charged residues occupy regions in or near to where similarly positively charged terminals of putrescine are expected to lie. A network connecting this region to the active site is also conserved in the human structure.	
<i>Plasmodium</i> -specific inserts	
Model predicts the position of three <i>Plasmodium</i> -specific inserts that possibly participate in interactions unique to the bifunctional complex. The shorter inserts are more conserved and appear to be more structurally defined.	

to severely compromise the enzyme. Therefore, it should now be possible to use the model to inform more refined deletion mutagenesis experiments. This would allow delineation of insert regions, which comprise the core folding region of AdoMetDC that is recognisable in other organisms, and regions that are *Plasmodium*-specific inserts. This strategy has already been successfully followed to demonstrate that some inserts are necessary for full functioning of the bifunctional enzyme (Birkholtz *et al.*, 2004).

From this study it was also possible to suggest which regions of the AdoMetDC domain are most likely to make contact with the rest of the bifunctional complex. This was based on the assumption that the more diverged regions would be required for novel interactions, and that the more conserved regions would fold into the recognised AdoMetDC core. The model could therefore also be used to guide experimental studies to determine the interacting regions of the bifunctional complex. Since the inserts represent a major point of divergence of the malarial enzyme from other organisms, it is expected that they mediate some of the interactions required for formation of the bifunctional protein. Since it appears that some of these regions are required for normal enzyme functioning, disrupting the folding of these regions may be a path to novel anti-malarials. Such disruption is potentially possible through the binding of small molecules or peptide mimics that can insinuate themselves into the bifunctional structure (Otvos *et al.*, 2000).

Of the inserts, two were relatively long and therefore no attempt was made to model these. Since no template was available, *ab initio* modelling would have to be relied upon. The computational difficulty of this for very long sequences precludes accurate *ab initio* structural modelling of these regions for the present. Some methods are emerging that make use of secondary structure predictions to construct sheet and helix elements, and then fold these in turn in order to obtain a tertiary structure (Godzik, 2003; Chivian *et al.*, 2003). These methods could be used in future to obtain tentative backbone traces of these regions in *P. falciparum* AdoMetDC. The structures of these regions are of interest since they represent



potential points of interference that are not present in the human host. It is presently unlikely that computer modelling will yield accurate enough information for the design of small molecules to do this. It is also unlikely that small molecules will be identified which will completely disrupt protein-protein interactions. The surface areas of protein-protein interactions tend to be too large ( $\pm 1200 \text{ \AA}^2$ ) for small molecules to be effective (C.A. Lipinski, personal communication). However, protein-protein interactions can be disrupted by small changes (e.g. point mutations) through transmitted effects (Myers *et al.*, 2001; Peterson and Schachman, 1992). Sequence analysis of these regions and *ab initio* modelling might also suggest possible peptide mimics for the disruption of protein-protein interactions, and may be another alternative to anti-malarials.

The problem of the inserts highlights the current state of knowledge regarding this enzyme. It is currently difficult to proceed further using *in silico* methods alone since much rests on knowing where the inserts begin and end. Therefore, experimental validation of the model is required before further progress in modelling can be made. Deletion mutagenesis as described is one technique that can be followed, since it is expected that insert regions will have less of an effect on enzyme functioning than if core regions of the protein were to be deleted. Since crystallisation of malarial proteins is difficult, X-ray crystallography is often precluded (Baca and Hol, 2000). There are a number of other techniques that could be followed to yield structural information. Yeast two-hybrid expression can be used to identify interacting proteins (Toby and Golemis, 2001), and combined with deletion studies could be used to identify interacting regions within the bifunctional complex. Electron microscopy and atomic force microscopy (AFM) could also potentially be used to gain insight into the macroscopic features of the bifunctional complex. Furthermore, fluid cell techniques in AFM could be used to determine protein-protein interactions (Fotiadis *et al.*, 2002; Alonso and Goldmann, 2003). Prior to this study inserts 1 and 2 were unidentified, therefore deletion mutagenesis of these inserts should be performed to determine their requirements for the bifunctional enzyme. The exact delineation of insert 3 and the hinge region is still uncertain, since they occur in regions of high sequence divergence. Therefore, further experimental studies need to be carried out in order to determine where the AdoMetDC core fold joins *Plasmodium*-specific regions. For this deletion mutagenesis is again suggested, followed by the determination on the effects of enzyme activity and protein-protein interactions. For exact delineation, incremental deletion from the predicted borders is suggested. Also, based on secondary structure predictions from the three known complete sequences for the bifunctional protein, predicted regions of  $\alpha$ -helical or  $\beta$ -sheet structure can be targeted, since disturbing these regions is expected to disrupt any larger folding that these regions may be undergoing. This may help determine to what extent these regions are forming defined folds, or whether they may be adopting a disordered structure.

The model revealed potential reasons for the lack of putrescine stimulation in the malarial enzyme, as opposed to the mammalian enzymes. A conserved network of charged residues (from the human and plant structures) that is thought to transmit the effects of putrescine binding to the active site, was observed in the model. At the putrescine binding site a number of interesting mutations were observed, compared to



the corresponding site in the human enzyme. A number of positively charged residues were observed in the regions where the amine terminals of putrescine would be expected to sit. It is suggested that some or all of these residues take over the functioning of putrescine. This was confirmed by experimental studies in which some of these residues were individually mutated in recombinantly expressed enzyme. From this Arg11 was found to be necessary for enzyme activity. It also appears that two Lys residues (Lys15 and Lys215) operate in tandem to simulate one end of putrescine. Similar studies suggest that internal residues perform the function of putrescine in plants and *Trypanosoma* (Clyne *et al.*, 2002; Bennett *et al.*, 2002). While this aspect of the study does not impact on the discovery of novel inhibitors, it does further overall understanding of the enzyme. It also contributes to the confidence of the overall accuracy of the model. Since the more diverged regions of the sequence are more difficult to model, experimental evidence can help to ascertain the model's accuracy. The region in the model that corresponds to the putrescine binding site in the human site is more diverged in sequence than the rest of the enzyme. The mutational studies suggest the model is correct in the putrescine associated regions and could be used for directing other experimental studies. Suggested further experiments would be to create a double mutant targeting the two lysine residues of this study (Lys15 and Lys215), to see whether a further reduction in activity would result. Also of interest is whether these residues and Arg11 have an effect on processing of the proenzyme. The near complete loss of activity for the Arg11Leu mutant may be due to an inability of the enzyme to self-cleave into the functional form. Stimulation of both processing and activity has so far only been observed in the mammalian enzyme (Stanley and Pegg, 1991; Xiong *et al.*, 1997). However, the effect of putrescine on processing of AdoMetDC in yeast – where activity is stimulated (Pösö *et al.*, 1975b; Hoyt *et al.*, 2000) – has yet to be determined. It would therefore appear that evolution of putrescine stimulation occurred at most twice, and that previous to this internal residues were required. It would also be interesting to see whether putrescine binding and stimulation could be engineered into the malarial enzyme, possibly in stages of an increasing need for the allosteric effector. Complementary studies could also be carried out in the human enzyme, to determine whether internal stimulation once existed. Such studies could yield insight into the evolution of putrescine stimulation for the mammalian class of enzymes, and possibly of allostery in general.

The active site of the model was fairly conserved when compared with the human enzyme. However, there were some substitutions as well as differences in the predicted shape of the active site. This suggested that it might be possible to identify inhibitors that are more effective against the parasite enzyme than the host enzyme. The model was used to screen against large libraries of small molecules to identify novel inhibitors. Some of the compounds identified from the NCI database were selected for biochemical testing on recombinantly expressed enzyme. The only criteria that was used to select compounds for testing was their predicted docking score. Biochemical testing unfortunately revealed none of these compounds to be likely inhibitors of malarial AdoMetDC. This was considered to be largely due to problems encountered with solubility. In future it is suggested, that compounds selected for testing should be subjected to other screens that predict solubility, lead-likeness and drug-likeness, etc (Egan and Lauri, 2002; Lipinski

*et al.*, 1997; Walters and Murcko, 2002). Combined use of a more than one docking algorithm for mass computational screening is also suggested in order to identify more likely inhibitors based on repeated hits (Charifson *et al.*, 1999). Modification of known inhibitors and/or substrates as scaffolds is also suggested, since this should increase the likelihood of identifying novel inhibitors. Specifically, the predicted cavity near the pyruvoyl residue, and the substitution of Thr245*hum* by a Ser residue may allow bulkier ligands to fit the *Plasmodium* enzyme. The substitution of Asn224*hum* with Thr416 may also allow for the design of ligands that exploit different polar interactions in the parasite. The model also suggests why Tris does not affect the *Plasmodium* enzyme (Section 2.4.3.2), due the replacement of His5*hum* by glycine. Tris inhibition could possibly be engineered into the *Plasmodium* enzyme to test this prediction, or conversely the human residue mutated to the corresponding *Plasmodium* residue.

The modelling of this enzyme highlighted the difficulties of modelling low homology proteins. During this study the need for integrating techniques when constructing a model was highlighted. In this case motif identification, inclusion of sequences of sister *Plasmodium* species and secondary structure all contributed to easing the difficulties of modelling such proteins. During this process some of the shortcomings of current *in silico* methods were also revealed, the lack of integration of existing methods chief among these. There is much information available when it comes to such projects, however, it is not being exploited to the fullest. For example no program could be found that can produce a multiple alignment by simultaneously using available sequences, secondary structures, known structures, known residue-residue contacts, etc. Whereas a number of methods exist to conduct such analysis individually much manual intervention is required to integrate all of these. If computational protein modelling is to become commonplace and reliable, however, automated integration of these methods will be needed or a vast improvement in the capabilities of *ab initio* modelling is required.

Nonetheless, following a computational approach it was possible to gather further insights into the structure of the bifunctional malarial ODC/AdoMetDC. Some of the predictions were confirmed experimentally. However, the objective of identifying novel inhibitors was not met. In order to fulfil this it is suggested that the different approaches described above should be followed. The use of computational methods for understanding and predicting protein structure and identifying novel inhibitors is becoming commonplace (Fauman *et al.*, 2003; Krumrine *et al.*, 2003). The rapid acquisition of resistance to existing drugs highlights the need for fast discovery of new drugs, and it is expected that a large array of methodologies will be required to fulfil this need. Computational drug discovery is but one technique that is likely to be required. Although computational drug design will probably be more difficult with malaria due to complications introduced by the uniqueness of it's genome and proteins, it is predicted that structural modelling will be indispensable in our fight against this parasitic diseases in the 21st century.



## Bibliography

- Albert, A., Dhanaraj, V., Genschel, U., Khan, G., Ramjee, M., Pulido, R., Sibanda, B., von Delft, F., Witty, M., Blundell, T., Smith, A., and Abell, C. (1998) Crystal structure of aspartate decarboxylase at 2.2 Å resolution provides evidence for an ester in protein self-processing. *Nat Struct Biol*, **5** (4), 289–93.
- Allen, R. R. and Klinman, J. P. (1981) Stereochemistry and kinetic isotope effects in the decarboxylation of S-adenosylmethionine as catalyzed by the pyruvoyl enzyme, S-adenosylmethionine decarboxylase. *J Biol Chem*, **256** (7), 3233–9.
- Almud, J., Oliveira, M., Kern, A., Grishin, N., Phillips, M., and Hackert, M. (2000) Crystal structure of human ornithine decarboxylase at 2.1 Å resolution: structural insights to antizyme binding. *J Mol Biol*, **295** (1), 7–16.
- Alonso, J. and Goldmann, W. (2003) Feeling the forces: atomic force microscopy in cell biology. *Life Sci*, **72** (23), 2553–60.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol*, **215** (3), 403–10.
- Anfinsen, C. (1973) Principles that govern the folding of protein chains. *Science*, **181** (96), 223–30.
- Assaraf, Y., Abu-Elheiga, L., Spira, D., Desser, H., and Bachrach, U. (1987) Effect of polyamine depletion on macromolecular synthesis of the malarial parasite, *Plasmodium falciparum*, cultured in human erythrocytes. *Biochem J*, **242** (1), 221–6.
- Ayala, F., Escalante, A., Altaf, A., and Rich, S. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 20. ASM Press, 285–300.
- Baca, A. and Hol, W. (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** (2), 113–8.
- Bacchi, C., Brun, R., Croft, S., Alicea, K., and Buhler, Y. (1996) In vivo trypanocidal activities of new S-adenosylmethionine decarboxylase inhibitors. *Antimicrob Agents Chemother*, **40** (6), 1448–53.
- Bailey, T. L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover

- motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28–36.
- Bennett, E. M., Ekstrom, J. L., Pegg, A. E., and Ealick, S. E. (2002) Monomeric S-adenosylmethionine decarboxylase from plants provides an alternative to putrescine stimulation. *Biochemistry*, **41** (49), 14509–17.
- Birkholtz, L., Joubert, F., Neitz, A., and Louw, A. (2003) Comparative properties of a three-dimensional model of *Plasmodium falciparum* ornithine decarboxylase. *Proteins*, **50** (3), 464–73.
- Birkholtz, L., Wrenger, C., Joubert, F., Wells, G., Walter, R., and Louw, A. (2004) Parasite-specific inserts in the bifunctional S-adenosylmethionine decarboxylase/ornithine decarboxylase of *Plasmodium falciparum* modulate catalytic activities and domain interactions. *Biochem J*, **377** (Pt 2), 439–48.
- Bitonti, A. J., McCann, P. P., and Sjoerdsma, A. (1987) *Plasmodium falciparum* and *Plasmodium berghei*: Effects of ornithine decarboxylase inhibitors on erythrocytic schizogony. *Exp Parasitol*, **64** (2), 237–43.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M., Estreicher, A., Gasteiger, E., Martin, M., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, **31** (1), 365–70.
- Böhm, H.-J. and Klebe, G. (1996) What can be learnt from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew Chem Int Ed Engl*, **35**, 2588–2614.
- Bradford, 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.
- Breman, J. (2001) The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg*, **64** (1-2 Suppl), 1–11.
- Brun, R., Buhler, Y., Sandmeier, U., Kaminsky, R., Bacchi, C., Rattendi, D., Lane, S., Croft, S., Snowdon, D., Yardley, V., Caravatti, G., Frei, J., Stanek, J., and Mett, H. (1996) In vitro trypanocidal activities of new S-adenosylmethionine decarboxylase inhibitors. *Antimicrob Agents Chemother*, **40** (6), 1442–7.
- Byers, T., Ganem, B., and Pegg, A. (1992) Cytostasis induced in L1210 murine leukaemia cells by the S-adenosyl-L-methionine decarboxylase inhibitor 5'-((Z)-4-amino-2-butenyl)methylamino)-5'-deoxyadenosine may be due to hypusine depletion. *Biochem J*, **287** ( Pt 3), 717–24.
- Byers, T. L., Wechter, R. S., Hu, R. H., and Pegg, A. E. (1994) Effects of the S-adenosylmethionine decarboxylase inhibitor, 5'-((z)-4-amino-2-butenyl)methylamino)-5'-deoxyadenosine, on cell growth and polyamine metabolism and transport in chinese hamster ovary cell cultures. *Biochem J*, **303** ( Pt 1),



89–96.

- Charifson, P., Corkery, J., Murcko, M., and Walters, W. (1999) Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem*, **42** (25), 5100–9.
- Chivian, D., Robertson, T., Bonneau, R., and Baker, D. (2003) *Ab initio* methods. *Methods Biochem Anal*, **44**, 547–57.
- Chothia, C. and Lesk, A. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J*, **5** (4), 823–6.
- Clarke, J., Scopes, D., Sodeinde, O., and Mason, P. (2001) Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. A novel bifunctional enzyme in malaria parasites. *Eur J Biochem*, **268** (7), 2013–9.
- Clyne, T., Kinch, L., and Phillips, M. (2002) Putrescine activation of *Trypanosoma cruzi* S-adenosylmethionine decarboxylase. *Biochemistry*, **41** (44), 13207–16.
- Coffino, P. (2000) Polyamines in spermiogenesis: not now, darling. *Proc Natl Acad Sci U S A*, **97** (9), 4421–3.
- Cowman, A. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 22. ASM Press, 317–330.
- Cramer, C. J. (2002) In *Essentials of Computational Chemistry*, Chapter 1-3. John Wiley & Sons, 1–92.
- Davis, A., Teague, S., and Kleywegt, G. (2003) Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew Chem Int Ed Engl*, **42** (24), 2718–36.
- Eckstein-Ludwig, U., Webb, R., Van Goethem, I., East, J., Lee, A., Kimura, M., O'Neill, P., Bray, P., Ward, S., and Krishna, S. (2003) Artemisinin target the SERCA of *Plasmodium falciparum*. *Nature*, **424** (6951), 957–61.
- Egan, W. and Lauri, G. (2002) Prediction of intestinal permeability. *Adv Drug Deliv Rev*, **54** (3), 273–89.
- Ekstrom, J. L., Mathews, I. I., Stanley, B. A., Pegg, A. E., and Ealick, S. E. (1999) The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 Å resolution reveals a novel fold. *Structure Fold Des*, **7** (5), 583–95.
- Ekstrom, J. L., Tolbert, W. D., Xiong, H., Pegg, A. E., and Ealick, S. E. (2001) Structure of a Human S-adenosylmethionine Decarboxylase Self-Processing Ester Intermediate and Mechanism of Putrescine Stimulation of Processing as Revealed by the H243A Mutant. *Biochemistry*, **40** (32), 9495–504.

- Fauman, E., Hopkins, A., and Groom, C. (2003) Structural bioinformatics in drug discovery. *Methods Biochem Anal*, **44**, 477–97.
- Fiser, A., Do, R. K., and Šali, A. (2000) Modeling of loops in protein structures. *Protein Sci*, **9** (9), 1753–73.
- Flower, D. R. (2002) Molecular Informatics: Sharpening Drug Design's Cutting Edge. *Drug Design Cutting edge approaches*, 1–53.
- Fotiadis, D., Scheuring, S., Müller, S., Engel, A., and Müller, D. (2002) Imaging and manipulation of biological structures with the AFM. *Micron*, **33** (4), 385–97.
- Gallagher, T., Rozwarski, D., Ernst, S., and Hackert, M. (1993) Refined structure of the pyruvoyl-dependent histidine decarboxylase from *Lactobacillus* 30a. *J Mol Biol*, **230** (2), 516–28.
- Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C., and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419** (6906), 498–511.
- Garfield, S. (2000) In *Mauve*. Faber & Faber.
- Gleeson, M. (2000) The plastid in Apicomplexa: what use is it? *Int J Parasitol*, **30** (10), 1053–70.
- Godzik, A. (2003) Fold recognition methods. *Methods Biochem Anal*, **44**, 525–46.
- Graminski, F., Carlson, C., Ziemer, J., Cai, F., Vermeulen, N., Vanderwerf, S., and Burns, M. (2002) Synthesis of Bis-spermine Dimers that are potent polyamine transport inhibitors. *Bioorg Med Chem Lett*, **12**, 35–40.
- Greenwood, B. and Mutabingwa, T. (2002) Malaria in 2002. *Nature*, **415** (6872), 670–2.
- Hafner, E., Tabor, C., and Tabor, H. (1979) Mutants of *Escherichia coli* that do not contain 1,4-diaminobutane (putrescine) or spermidine. *J Biol Chem*, **254** (24), 12419–26.
- Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, **47** (4), 409–43.
- Hargreaves, K., Koekemoer, L., Brooke, B., Hunt, R., Mthembu, J., and Coetzee, M. (2000) *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Med Vet Entomol*, **14** (2), 181–9.



- Hoffman, S., Subramanian, G., Collins, F., and Venter, J. (2002) *Plasmodium*, human and *Anopheles* genomics and malaria. *Nature*, **415** (6872), 702–9.
- Hoyt, M. A., Williams-Abbott, L. J., Pitkin, J. W., and Davis, R. H. (2000) Cloning and expression of the S-adenosylmethionine decarboxylase gene of *Neurospora crassa* and processing of its product. *Mol Gen Genet*, **263** (4), 664–73.
- Hyde, J., Kelly, S., Holloway, S., Snewin, V., and Sims, P. (1989) A general approach to isolating *Plasmodium falciparum* genes using non-redundant oligonucleotides inferred from protein sequences of other organisms. *Mol Biochem Parasitol*, **32** (2-3), 247–61.
- Igarashi, K., Sakamoto, I., Goto, N., Kashiwagi, K., Honma, R., and Hirose, S. (1982) Interaction between polyamines and nucleic acids or phospholipids. *Arch Biochem Biophys*, **219** (2), 438–43.
- Jensen, F. (1999) In *Introduction to computational chemistry*, Chapter 14. John Wiley & Sons, 316–346.
- Joubert, F., Neitz, A., and Louw, A. (2001) Structure-based inhibitor screening: a family of sulfonated dye inhibitors for malaria parasite triosephosphate isomerase. *Proteins*, **45** (2), 136–43.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22** (12), 2577–637.
- Kinch, L. N., Scott, J. R., Ullman, B., and Phillips, M. A. (1999) Cloning and kinetic characterization of the *Trypanosoma cruzi*. *Mol Biochem Parasitol*, **101** (1-2), 1–11.
- Krieger, E., Nabuurs, S., and Vriend, G. (2003) Homology modeling. *Methods Biochem Anal*, **44**, 509–23.
- Krogstad, D. and Dibyendu, D. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 23. ASM Press, 331–340.
- Krumrine, J., Raubacher, F., Brooijmans, N., and Kuntz, I. (2003) Principles and methods of docking and ligand design. *Methods Biochem Anal*, **44**, 443–76.
- Lipinski, C., Lombardo, F., Dominy, B., and Feeney, P. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **23**, 3–25.
- Liu, H., Elstner, M., Kaxiras, E., Frauenheim, T., Hermans, J., and Yang, W. (2001) Quantum mechanics simulation of protein dynamics on long timescale. *Proteins*, **44** (4), 484–9.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000) comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*, **29**, 291–325.
- Marton, L. J. and Pegg, A. E. (1995) Polyamines as targets for therapeutic intervention. *Annu Rev*

*Pharmacol Toxicol*, **35**, 55–91.

McKie, J., Douglas, K., Chan, C., Roser, S., Yates, R., Read, M., Hyde, J., Dascombe, M., Yuthavong, Y., and Sirawaraporn, W. (1998) Rational drug design approach for overcoming drug resistance: application to pyrimethamine resistance in malaria. *J Med Chem*, **41** (9), 1367–70.

Meshnick, S. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 24. ASM Press, 341–354.

Milhous, W. and Dennis, D. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 21. ASM Press, 303–316.

Miller, L., Baruch, D., Marsh, K., and Doumbo, O. (2002) The pathogenic basis of malaria. *Nature*, **415** (6872), 673–9.

Minton, A. (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem*, **276** (14), 10577–80.

Mitchell, J. and Rusch, H. (1973) Regulation of polyamine synthesis in *Physarum polycephalum* during growth and differentiation. *Biochim Biophys Acta*, **297**, 503–516.

Morris, A., MacArthur, M., Hutchinson, E., and Thornton, J. (1992) Stereochemical quality of protein structure coordinates. *Proteins*, **12** (4), 345–64.

Myers, D., Jackson, L., Ipe, V., Murphy, G., and Phillips, M. (2001) Long-range interactions in the dimer interface of ornithine decarboxylase are important for enzyme function. *Biochemistry*, **40** (44), 13230–6.

Müller, S., Coombs, G. H., and Walter, R. D. (2001) targeting polyamines of parasitic protozoa in chemotherapy. *Trends Parasitol*, **17** (5), 242–9.

Müller, S., Da'dara, A., Lüersen, K., Wrenger, C., Das Gupta, R., Madhubala, R., and Walter, R. D. (2000) In the human malaria parasite *Plasmodium falciparum*, polyamines are synthesized by a bifunctional ornithine decarboxylase, S-adenosylmethionine decarboxylase. *J Biol Chem*, **275** (11), 8097–102.

Müller, S., Liebau, E., Walter, R., and Krauth-Siegel, R. (2003) Thiol-based redox metabolism of protozoan parasites. *Trends Parasitol*, **19** (7), 320–8.

Otvos, Jr, L., O, I., Rogers, M., Consolvo, P., Condie, B., Lovas, S., Bulet, P., and Blaszczyk-Thurin, M. (2000) Interaction between heat shock proteins and antimicrobial peptides. *Biochemistry*, **39** (46), 14150–9.



- Pankaskie, M. and Abdel-Monem, M. M. (1980) Inhibitors of polyamine biosynthesis. 8. Irreversible inhibition of mammalian S-adenosyl-L-methionine decarboxylase by substrate analogues. *J Med Chem*, **23** (2), 121–7.
- Pegg, A. and Williams-Ashman, H. (1969) On the role of S-adenosylmethionine in the Biosynthesis of Spermidine by Rat Prostate. *J Biol Chem*, **244**, 682–693.
- Pegg, A. E. (1984) S-adenosylmethionine decarboxylase: a brief review. *Cell Biochem Funct*, **2** (1), 11–5.
- Pegg, A. E. and Jacobs, G. (1983) Comparison of inhibitors of S-adenosylmethionine decarboxylase from different species. *Biochem J*, **213** (2), 495–502.
- Persson, K., Åslund, L., Grahn, B., Hanke, J., and Heby, O. (1998) *Trypanosoma cruzi* has not lost its S-adenosylmethionine decarboxylase: characterization of the gene and the encoded enzyme. *Biochem J*, **333** ( Pt 3), 527–37.
- Peterson, C. and Schachman, H. (1992) Long range effects of amino acid substitutions in the catalytic chain of aspartate transcarbamoylase. Localized replacements in the carboxyl-terminal alpha-helix cause marked alterations in allosteric properties and intersubunit interactions. *J Biol Chem*, **267** (4), 2443–50.
- Pizzi, E. and Frontali, C. (2001) Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res*, **11** (2), 218–29.
- Poroikov, V. V., Filimonov, D. A., Ihlenfeldt, W. D., Glorizova, T. A., Lagunin, A. A., Borodina, Y. V., Stepanchikova, A. V., and Nicklaus, M. C. (2003) PASS Biological Activity Spectrum Predictions in the Enhanced Open NCI Database Browser. *J Chem Inf Comput Sci*, **43**, 228–236.
- Pösö, H., Sinervirta, R., Himberg, J., and Jänne, J. (1975a) Putrescine-insensitive S-adenosyl-L-methionine decarboxylase from *Tetrahymena pyriformis*. *Acta Chem Scand B*, **29** (9), 932–6.
- Pösö, H., Sinervirta, R., and Jänne, J. (1975b) S-adenosylmethionine decarboxylase from baker's yeast. *Biochem J*, **151** (1), 67–73.
- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol*, **7**, 95–99.
- Rastelli, G., Sirawaraporn, W., Sompornpisut, P., Vilaivan, T., Kamchonwongpaisan, S., Quarrell, R., Lowe, G., Thebtaranonth, Y., and Yuthavong, Y. (2000) Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with *Plasmodium falciparum* dihydrofolate reductase: structural basis of antifolate resistance. *Bioorg Med Chem*, **8** (5), 1117–28.

- Recsei, P. A. and Snell, E. E. (1984) Pyruvate enzymes. *Annu Rev Biochem*, **53**, 357–87.
- Ridley, R. (2002) Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature*, **415** (6872), 686–93.
- Rieder, M. J., Taylor, S. L., Tobe, V. O., and Nickerson, D. A. (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res*, **26** (4), 967–73.
- Rishton, G. (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov Today*, **8** (2), 86–96.
- Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng*, **12** (2), 85–94.
- Salcedo, E., Cortese, J., Plowe, C., Sims, P., and Hyde, J. (2001) A bifunctional dihydrofolate synthetase–folypolyglutamate synthetase in *Plasmodium falciparum* identified by functional complementation in yeast and bacteria. *Mol Biochem Parasitol*, **112** (2), 239–52.
- Šali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, **234** (3), 779–815.
- Sander, C. and Schneider, R. (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9** (1), 56–68.
- Schwede, T., Kopp, J., Guex, N., and Peitsch, M. (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, **31** (13), 3381–5.
- Seely, J., Pösö, H., and Pegg, A. (1982) Purification of ornithine decarboxylase from kidneys of androgen-treated mice. *Biochemistry*, **21** (14), 3394–9.
- Sherman, I. W. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 1. ASM Press, 3–10.
- Shoichet, B., McGovern, S., Wei, B., and Irwin, J. (2002) Lead discovery using molecular docking. *Curr Opin Chem Biol*, **6** (4), 439–46.
- Singer, G. and Hickey, D. (2000) Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol*, **17** (11), 1581–8.
- Stanley, B. A. and Pegg, A. E. (1991) Amino acid residues necessary for putrescine stimulation of human S-adenosylmethionine decarboxylase proenzyme processing and catalytic activity. *J Biol Chem*, **266** (28), 18502–6.
- Stanley, B. A., Pegg, A. E., and Holm, I. (1989) Site of pyruvate formation and processing of mammalian



- S-adenosylmethionine decarboxylase proenzyme. *J Biol Chem*, **264** (35), 21073–9.
- Stanley, B. A., Shantz, L. M., and Pegg, A. E. (1994) Expression of mammalian S-adenosylmethionine decarboxylase in *Escherichia coli*. determination of sites for putrescine activation of activity and processing. *J Biol Chem*, **269** (11), 7901–7.
- Stoesser, G., Baker, W., van den Broek, A., Garcia-Pastor, M., Kanz, C., Kulikova, T., Leinonen, R., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., Nardone, F., Stoehr, P., Tuli, M., Tzouvara, K., and Vaughan, R. (2003) The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res*, **31** (1), 17–22.
- Tabor, C. and Tabor, H. (1985) Polyamines in microorganisms. *Microbiol Rev*, **49** (1), 81–99.
- Tabor, C. W. and Tabor, H. (1984a) Methionine adenosyltransferase (S-adenosylmethionine synthetase) and S-adenosylmethionine decarboxylase. *Adv Enzymol Relat Areas Mol Biol*, **56**, 251–82.
- Tabor, C. W. and Tabor, H. (1984b) Polyamines. *Annu Rev Biochem*, **53**, 749–90.
- Taylor, R., Jewsbury, P., and Essex, J. (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des*, **16** (3), 151–66.
- The *Plasmodium* Genome Database Collaborative (2001) Plasmodb: an integrative database of the *Plasmodium falciparum* genome. tools for accessing and analyzing finished and unfinished sequence data. *Nucleic Acids Res*, **29** (1), 66–69.
- The PDB Team (2003) The Protein Data Bank. *Methods Biochem Anal*, **44**, 181–98.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, **25** (24), 4876–82.
- Toby, G. and Golemis, E. (2001) Using the yeast interaction trap and other two-hybrid-based approaches to study protein-protein interactions. *Methods*, **24** (3), 201–17.
- Tolbert, W., Graham, D., White, R., and Ealick, S. (2003a) Pyruvoyl-dependent arginine decarboxylase from *Methanococcus jannaschii*: crystal structures of the self-cleaved and S53A proenzyme forms. *Structure Camb*, **11** (3), 285–94.
- Tolbert, W. D., Ekstrom, J. L., Mathews, I. I., Kapoor, P., Pegg, A. E., and Ealick, S. E. (2001) The Structural Basis for Substrate Specificity and Inhibition of Human S-adenosylmethionine Decarboxylase. *Biochemistry*, **40** (32), 9484–94.
- Tolbert, W. D., Zhang, Y., Cottet, S. E., Bennett, E. M., Ekstrom, J. L., Pegg, A. E., and Ealick, S. E.

- (2003b) Mechanism of human S-adenosylmethionine decarboxylase proenzyme processing as revealed by the structure of the S68A mutant. *Biochemistry*, **42** (8), 2386–95.
- Vaidya, A. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 25. ASM Press, 355–368.
- van den Berg, B., Ellis, R., and Dobson, C. (1999) Effects of macromolecular crowding on protein folding and aggregation. *EMBO J*, **18** (24), 6927–33.
- Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) Ligplot: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, **8** (2), 127–34.
- Walters, W. and Murcko, M. (2002) Prediction of 'drug-likeness'. *Adv Drug Deliv Rev*, **54** (3), 255–71.
- Wang, C. C. (1995) molecular mechanisms and therapeutic approaches to the treatment of African trypanosomiasis. *Annu Rev Pharmacol Toxicol*, **35**, 93–127.
- White, N. (1998) In *Malaria: Parasite Biology, Pathogenesis and Protection*, Chapter 26. ASM Press, 371–386.
- Withers-Martinez, C., Carpenter, E., Hackett, F., Ely, B., Sajid, M., Grainger, M., and Blackman, M. (1999) PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng*, **12** (12), 1113–20.
- Wrenger, C., Lüersen, K., Krause, T., Müller, S., and Walter, R. D. (2001) the emPlasmodium falciparumem bifunctional ornithine decarboxylase,S-adenosyl-l-methionine decarboxylase, enables a well balanced polyamine synthesis without domain-domain interaction. *J Biol Chem*, **276** (32), 29651–6.
- Xiong, H. and Pegg, A. (1999) Mechanistic studies of the processing of human S-adenosylmethionine decarboxylase proenzyme. Isolation of an ester intermediate. *J Biol Chem*, **274** (49), 35059–66.
- Xiong, H., Stanley, B. A., and Pegg, A. E. (1999) Role of Cysteine-82 in the Catalytic Mechanism of Human S-adenosylmethionine Decarboxylase. *Biochemistry*, **38** (8), 2462–70.
- Xiong, H., Stanley, B. A., Tekwani, B. L., and Pegg, A. E. (1997) Processing of mammalian and plant S-adenosylmethionine decarboxylase proenzymes. *J Biol Chem*, **272** (45), 28342–8.
- Xue, H. and Forsdyke, D. (2003) Low-complexity segments in *Plasmodium falciparum* proteins are primarily nucleic acid level adaptations. *Mol Biochem Parasitol*, **128** (1), 21–32.
- Yuthavong, Y., Vilaivan, T., Chareonsethakul, N., Kamchonwongpaisan, S., Sirawaraporn, W., Quarrell, R., and Lowe, G. (2000) Development of a lead inhibitor for the A16V+S108T mutant of dihydrofolate reductase from the cycloguanil-resistant strain (T9/94) of *Plasmodium falciparum*. *J Med Chem*,



43 (14), 2738–44.

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M., and Yuthavong, Y. (2003) Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol*, **10** (5), 357–65.

## Appendix A

### Supplementary data for chapter 2

#### CLUSTALX protein colouring:

Green	Thr, Ser, Gln, Asn
Cyan	Ala, Val, Ile, Leu, Met, Phe, Trp
Blue	Tyr, His
Magenta	Asp, Glu
Yellow	Pro
Orange	Gly
Pink	Cys, Lys, Arg

#### Swissprot accession numbers for multiple sequence alignment

<i>Bos taurus</i>	P50243
<i>Homo sapiens</i>	P17707, Q9BWK4
<i>Mesocricetus auratus</i>	P28918
<i>Mus musculus</i>	P31154
<i>Rattus norvegicus</i>	P17708
<i>Xenopus laevis</i>	P79888
<i>Drosophila melanogaster</i>	P91931, P91925, Q9VKY9
<i>Caenorhabditis elegans</i>	O02655
<i>Onchocerca volvulus</i>	Q27883
<i>Leishmania donovani</i>	Q25264
<i>Trypanosoma brucei brucei</i>	P50244
<i>Trypanosoma cruzi</i>	O76240, Q9UAD2
<i>Arabidopsis thaliana</i>	Q96286, Q96531, Q9M893
<i>Brassica juncea</i>	Q42613
<i>Catharanthus roseus</i>	Q42679



<i>Datura stramonium</i>	Q96555
<i>Dianthus caryophyllus</i>	Q39676
<i>Helianthus annuus</i>	O65354
<i>Hordeum chilense</i>	Q42829
<i>Zea mays</i>	O24575
<i>Nicotiana sylvestris</i>	O80402
<i>Oryza sativa</i>	O24215, O81269
<i>Pisum sativum</i>	Q43820
<i>Pharbitis nil</i>	Q96471
<i>Solanum tuberosum</i>	Q04694
<i>Spinacia oleracea</i>	P46255
<i>Nicotiana tabacum</i>	O04009, O49005
<i>Saccharomyces cerevisiae</i>	P21182





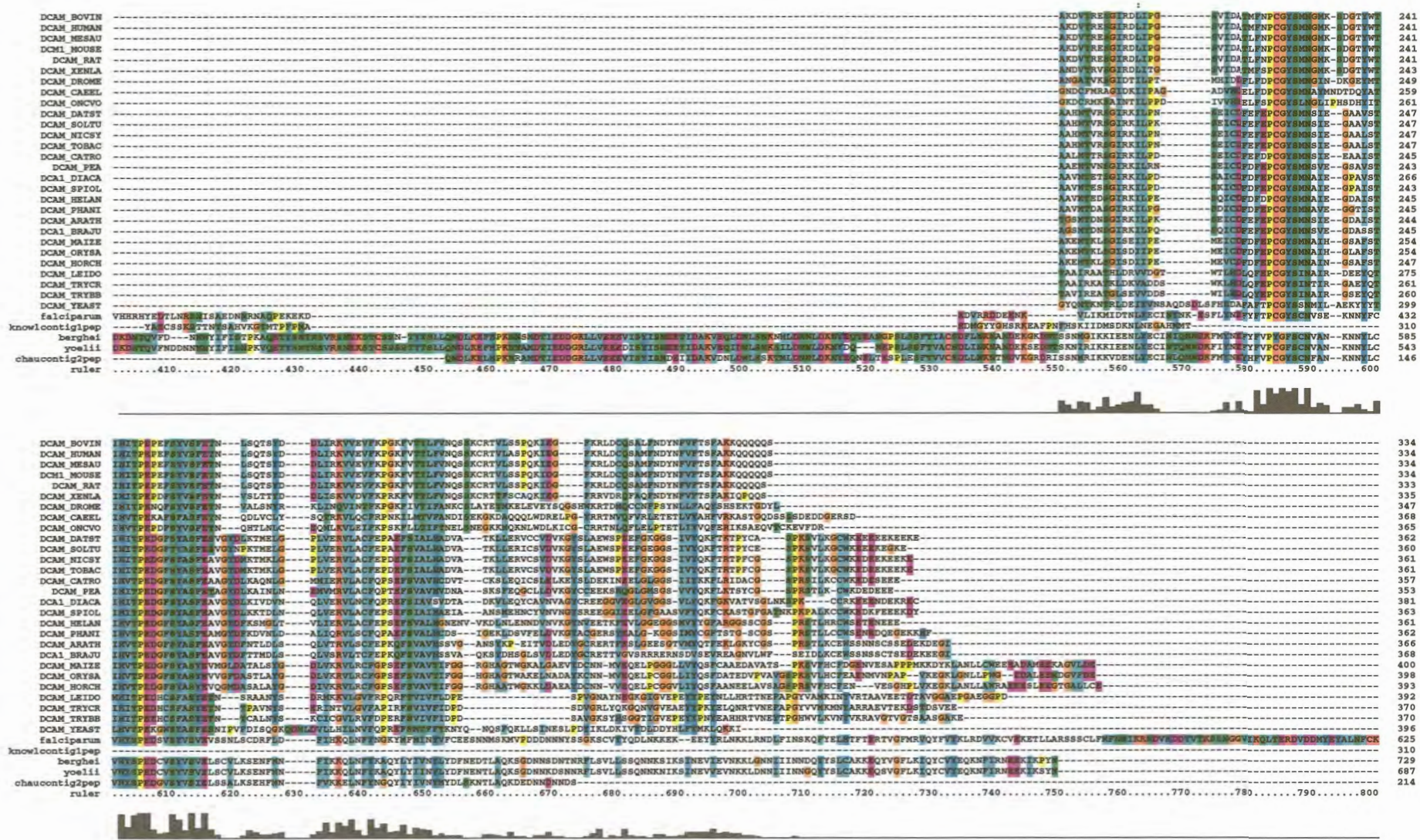


Figure A.1: Multipe alignment continued.



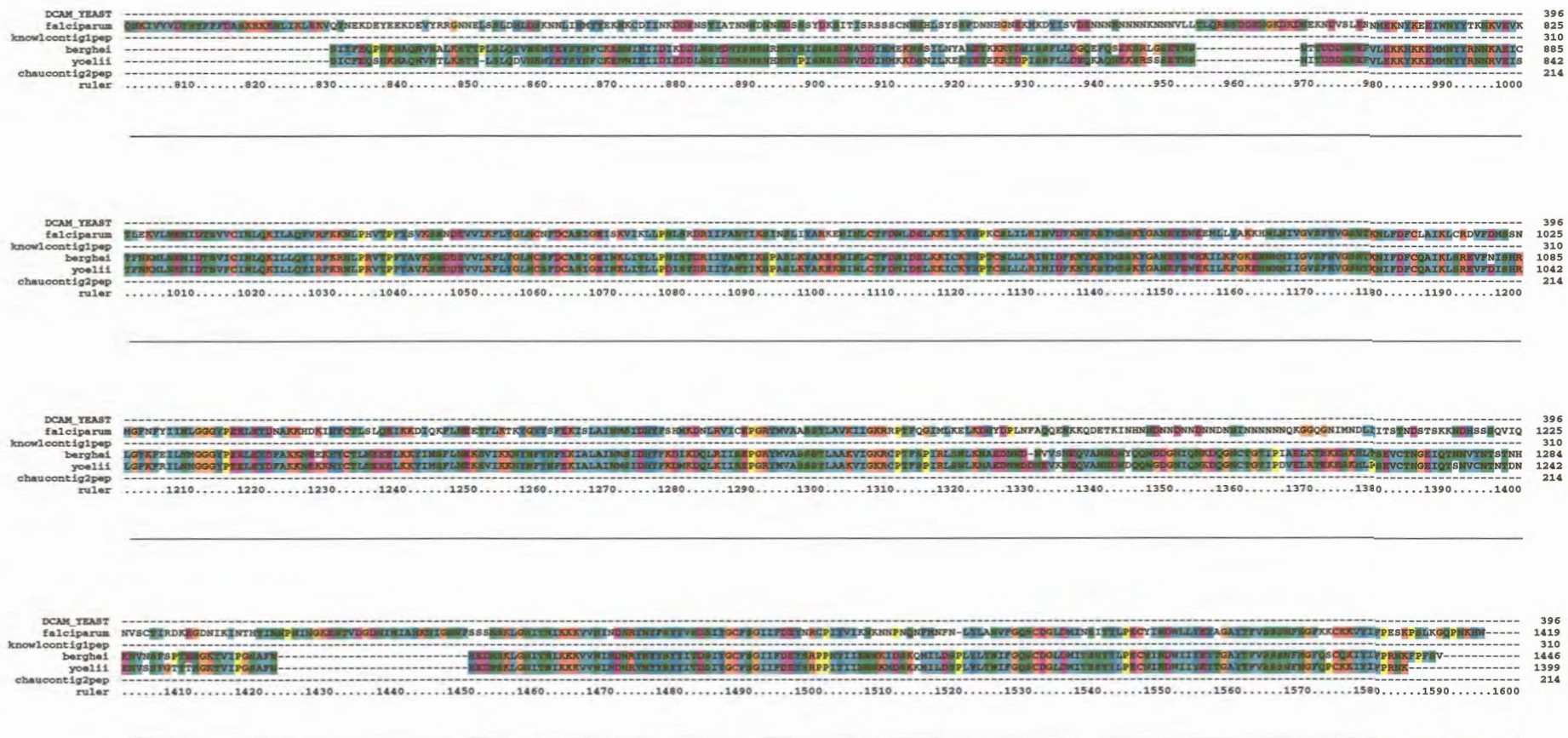


Figure A.1: Multiple alignment concluded.



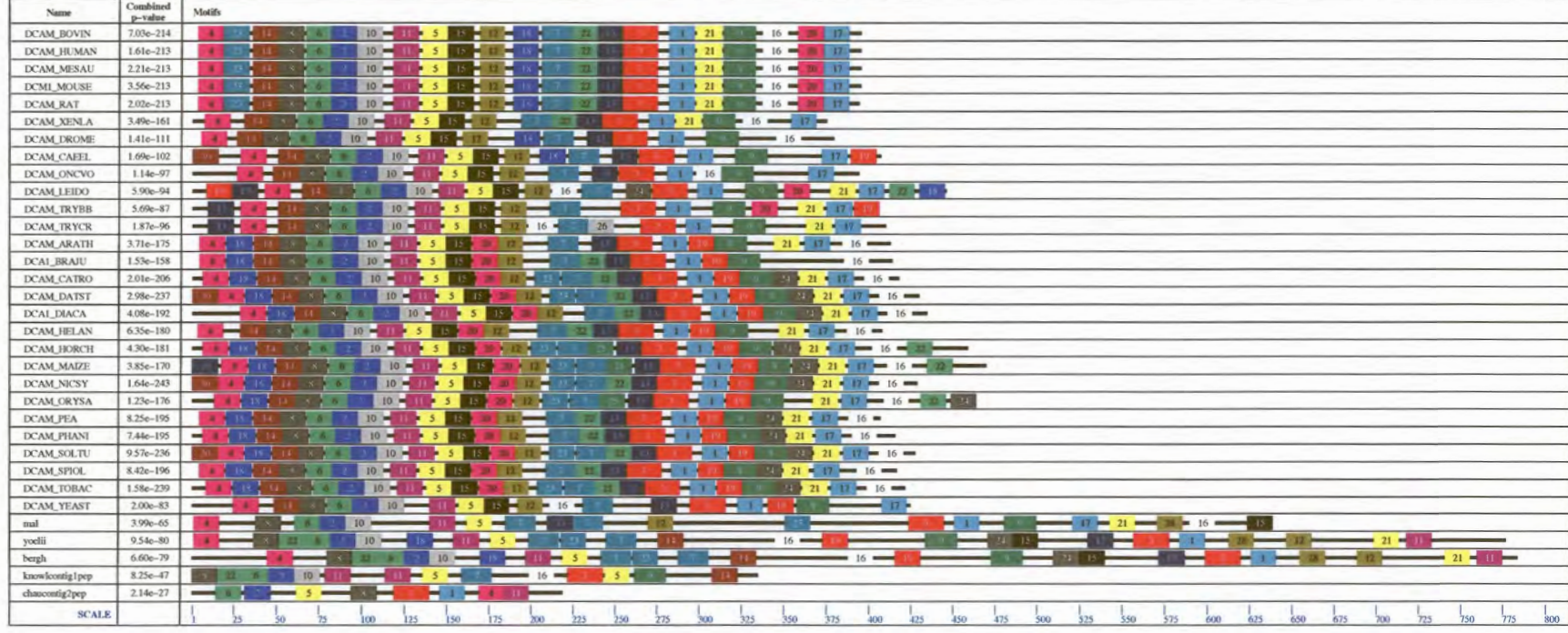


Figure A.2: Meme motifs of the set of unaligned sequences. Homologous motifs are numbered and coloured the same. Higher numbers indicate greater conservation. Motif 5 which roughly corresponds to helix 6 of the human enzyme was used to adjust the alignment.

## Appendix B

## Supplementary data for chapter 4

Table B.1: Hits identified from virtual screening against the internal LUDI BIOSYM database

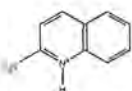
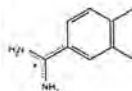
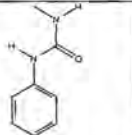
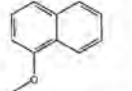
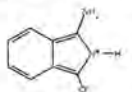
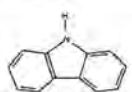
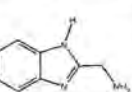
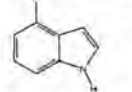
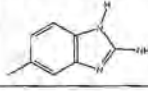
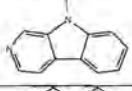
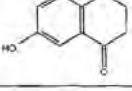
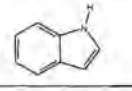
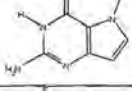
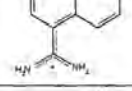
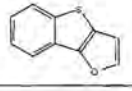
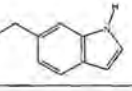
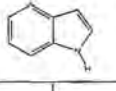
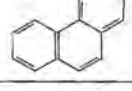
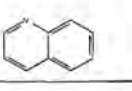
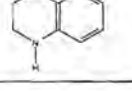

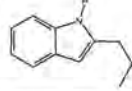
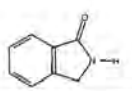
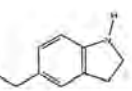
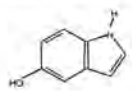
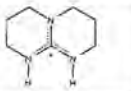
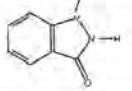
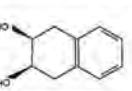
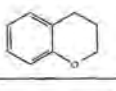
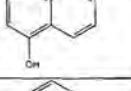
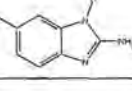
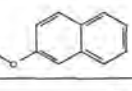
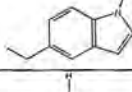
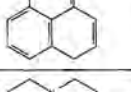
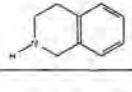
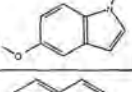
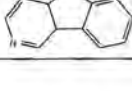
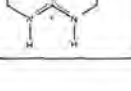
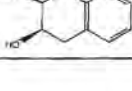
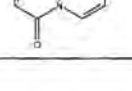
Human	Score	Model	Score	Human	Score	Model	Score
	638		680		524		549
	614		676		521		545
	585		667		518		538
	577		658		505		530
	563		625		503		524
	547		581		500		519
	544		574		500		515
	542		563		498		512
	534		563		487		508
	530		561		486		508



Table B.2: Hits identified from virtual screening against the ACD database

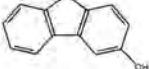
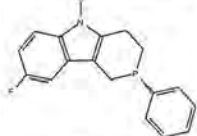
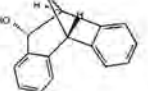
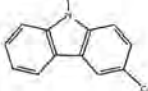

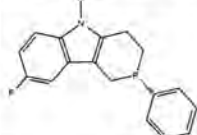
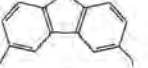
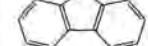
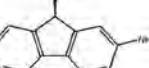
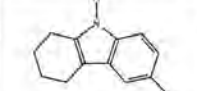
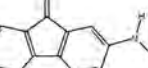
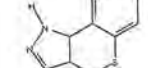
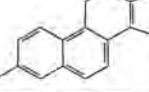
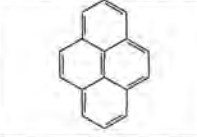
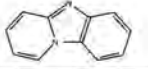

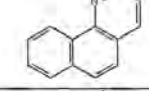
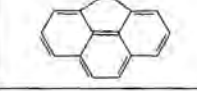
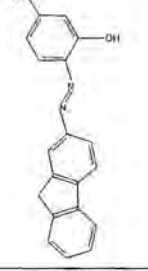
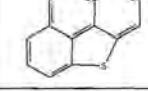
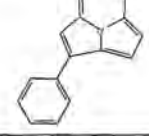
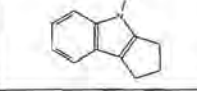
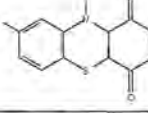
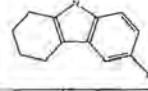
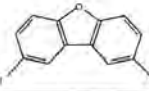

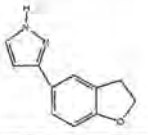

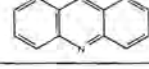

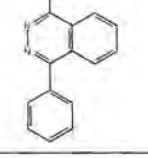
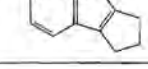
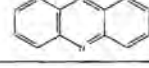
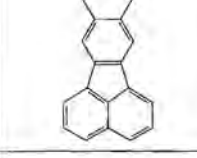


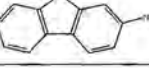
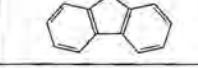
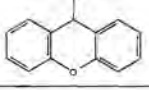

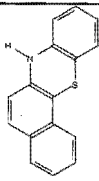
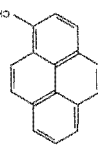
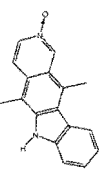
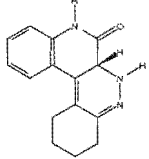
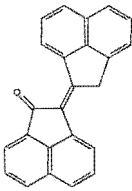
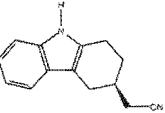
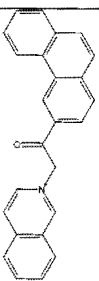
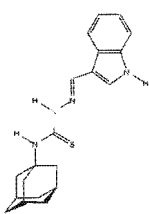
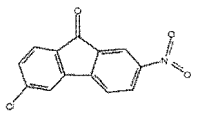
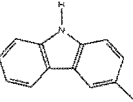
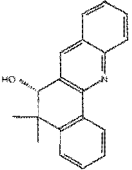
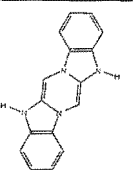
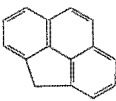
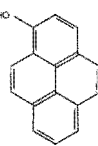
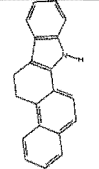
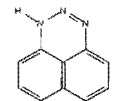
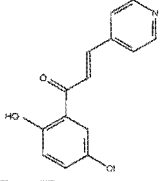
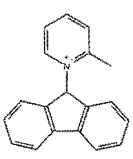
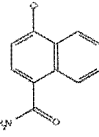
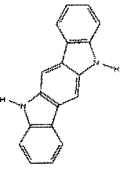
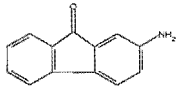
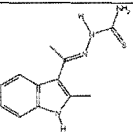
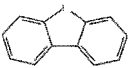
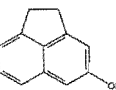
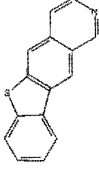
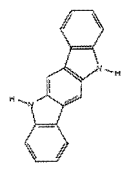
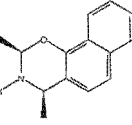
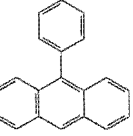
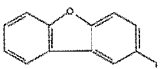
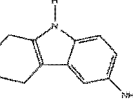
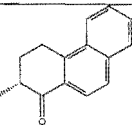
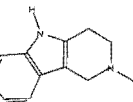
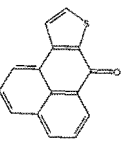
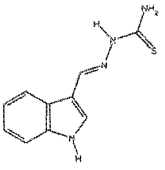
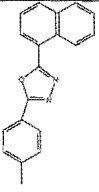
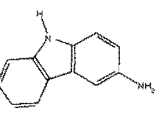
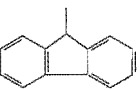
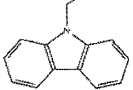
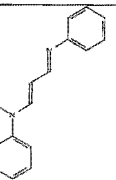
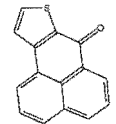
Human	Score	Model	Score	Human	Score	Model	Score
	718		832		674		709
	711		817		670		702
	709		747		668		698
	701		731		658		693
	698		729		655		692
	692		725		648		689
	688		722		646		687
	685		722		645		684
	683		714		643		676
	683		710		643		676

Table B.3: Hits identified from virtual screening against the NCI database

Human - NCI	Score	Model - NCI	Score	Human - NCI	Score	Model - NCI	Score
	717		832		649		709
	701		817		645		702
	690		747		645		698
	673		731		643		693
	672		729		641		692
	665		725		636		689
	661		722		636		687
	658		722		635		684
	656		714		635		676
	655		710		628		676