Chapter 1

Introduction

1.1. The need for new anti-malarials

Malaria is a disease characterised particularly by severe febrility. The disease acquired it's 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. It's value had a number

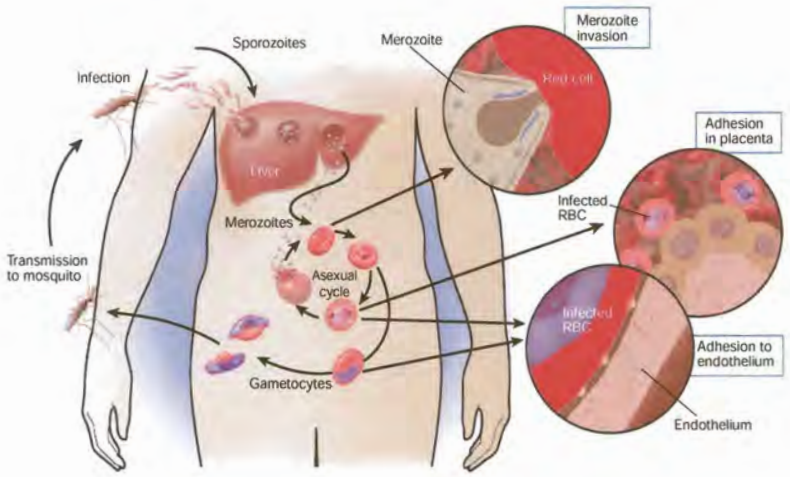

Figure 1.1: The life cycle of *P. falciparum*. Disease results from the asexual/red blood cell stage of the life cycle.

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 (Dichlorodiphenyltrichloroethane) 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 uspread 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 it's pharmacological profile a number of derivatives have been developed. This drug has been recently identified to target a 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 (Breman, 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 doxycyclin, 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).

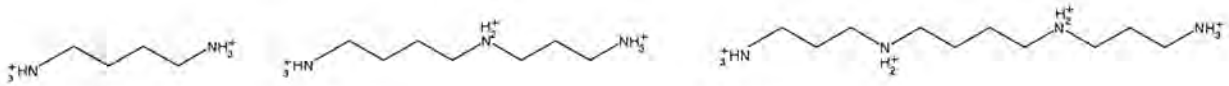

Figure 1.3: The main polyamines.

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 postranslational 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^1-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 proteosome for degradation. Antizyme acts catalytically in this regard, and can be recycled for degradation of more than

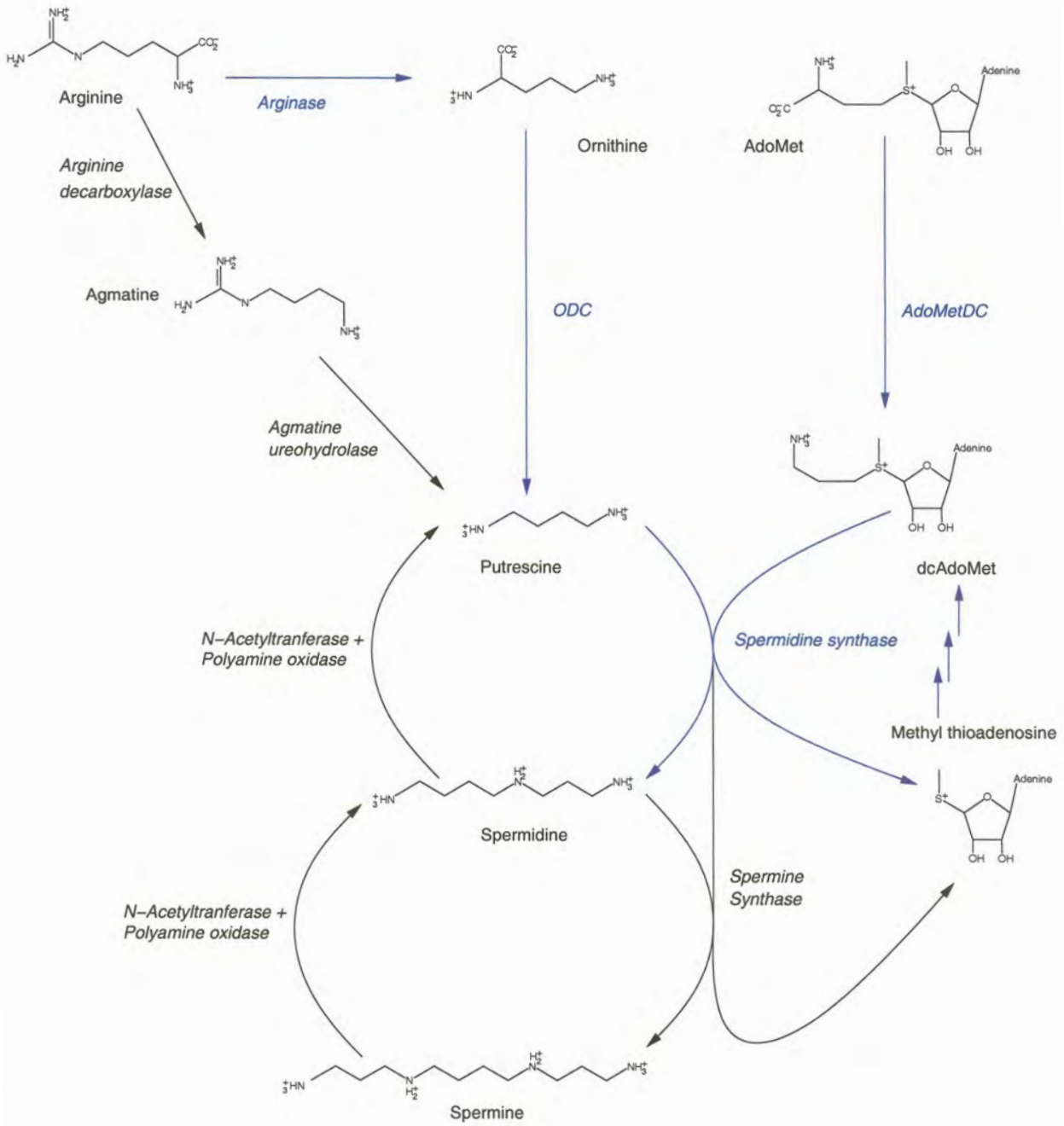

Figure 1.4: The generic polyamine pathway based on the mammalian system. Pathways and enzymes coloured blue indicate those that have been found in *P. falciparum*. ODC: Ornithine decarboxylase, AdoMetDC: S-adenosylmethionine decarboxylase.

one ODC molecule. This is in contrast to most proteins which are targeted for proteosome 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 (±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 cytoxicity (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 α-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 ± 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 α-chain (38.4 kDa) and the smaller N-terminal β-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 oxyoxazolidine intermediate (Fig. 1.5). This undergoes an N→O acyl shift to form an ester intermediate, β-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 α-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 α-carbon during the β-elimination step (Tolbert *et al.*, 2001). A summary of important mutations and their effects is given in Table 1.1.

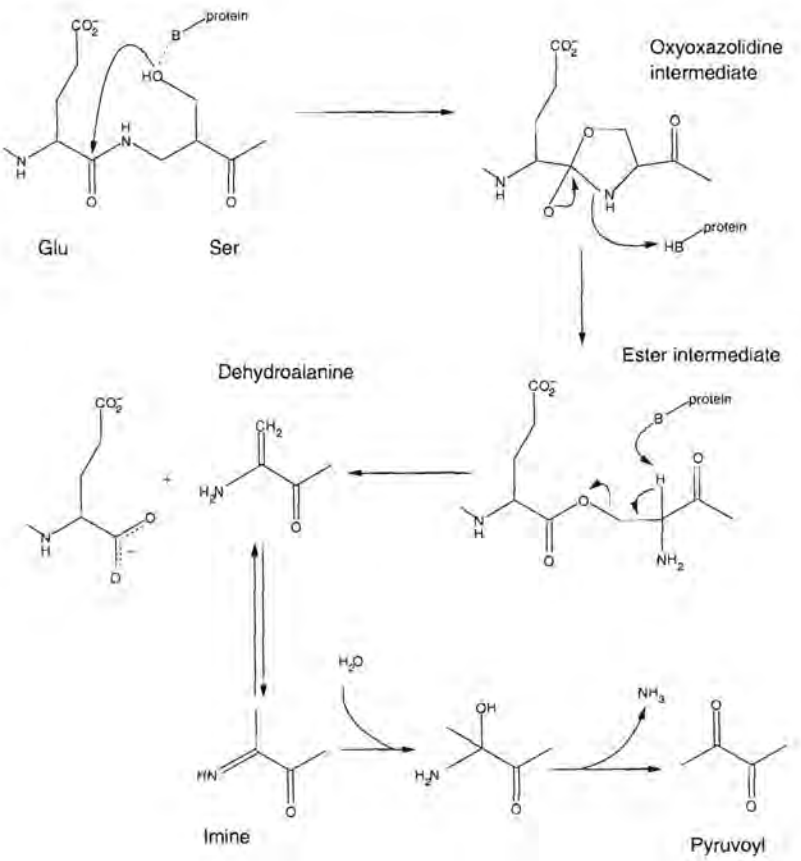

Figure 1.5: Formation of the AdoMetDC pyruvoyl residue (Bennett *et al.*, 2002).

Table 1.1: Effects of key mutations in human AdoMetDC.

Residue	Mutant	Processing	Activity	Function	Ref.
Ser68	Ala	Eliminated	Eliminated	Converted to pyruvate	(Stanley *et al.*, 1989) (Tolbert *et al.*, 2003b)
	Cys	Slowed	Converted to thiocarboxylate		(Xiong *et al.*, 1997)
	Thr	Slowed	Converted to α-ketobutyrate		(Xiong *et al.*, 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 *et al.*, 2003b)
	Ser	Slower	Not specified		(Tolbert *et al.*, 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 *et al.*, 1999)
	Lys	Eliminated	Eliminated		(Stanley *et al.*, 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 *et al.*, 1994)
Glu256	Gln	Not stimulated by putrescine	Substantially reduced Not stimulated by putrescine	Required for putrescine stimulation of processing and activity	(Stanley *et al.*, 1994)
Tyr112	Ala	Eliminated	Eliminated		(Stanley *et al.*, 1994)
Glu133	Gln	Putrescine absolutely required for processing	Almost eliminated		(Stanley *et al.*, 1994)
Asp174	Asn	Not stimulated by putrescine	Not stimulated by putrescine	Required for putrescine stimulation of processing and activity	(Xiong *et al.*, 1997)
His243	Ala	Incomplete processing Enzyme trapped in ester intermediate	Eliminated	Proton abstraction during β-elimination of ester intermediate	(Xiong *et al.*, 1999) (Ekstrom *et al.*, 2001)
	Glu	Slower	Eliminated		(Xiong *et al.*, 1999)
Ser229	Ala	Eliminated	Eliminated	Possible proton donor during processing	(Xiong and Pegg, 1999) (Tolbert *et al.*, 2003b)
	Cys	Very slow	Eliminated	May be required for product release during catalysis	(Xiong and Pegg, 1999) (Tolbert *et al.*, 2003b)
	Thr	Normal	Eliminated	May increase nucleophilicity of carbonyl carbon during processing	(Xiong and Pegg, 1999) (Tolbert *et al.*, 2003b)
Phe7	Ala	Normal	Substantially reduced	Required for correct binding of substrates and inhibitors	(Tolbert *et al.*, 2001)
Phe223	Ala	Normal	Substantially reduced	Required for correct binding of substrates and inhibitors	(Tolbert *et al.*, 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-$C_\alpha$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 reprotonated 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).

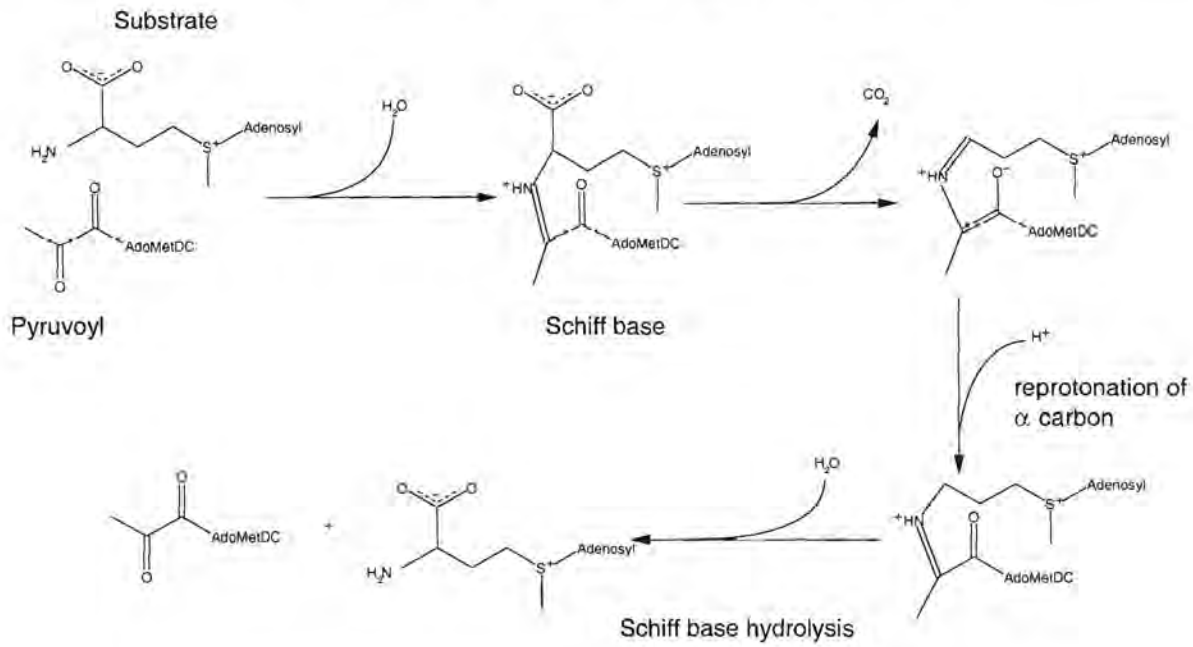

Figure 1.6: AdoMetDC reaction mechanism (Tolbert *et al.*, 2001).

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 β-sheets stacked between two layers of α-helices. The active site lies between the two β-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 β-strands B7 and B15 from each monomer (Fig. 1.8). The potato enzyme does not dimerise (Bennett *et al.*, 2002).

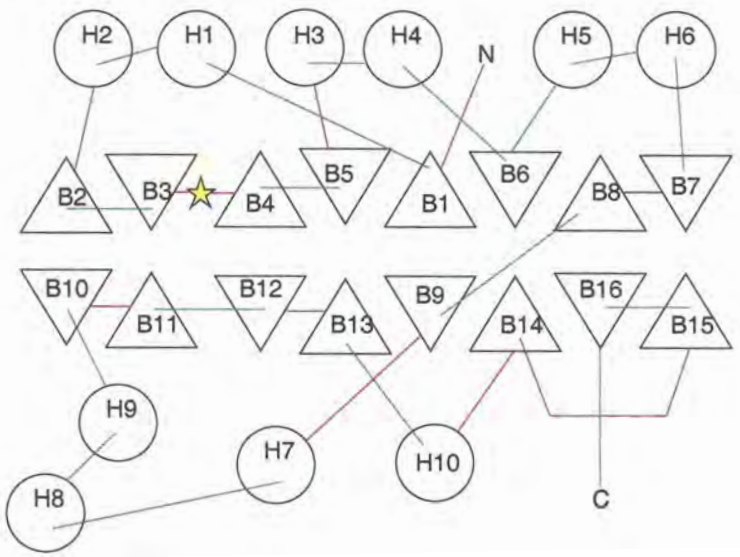

Figure 1.7: Topology of human AdoMetDC. Circles represent α-helices, triangles β-sheets. The active site and site of proteolytic cleavage is designated by the yellow star. N: N-terminus, C: C-terminus. Anterior loops are coloured green, and posterior loops magenta.

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-*

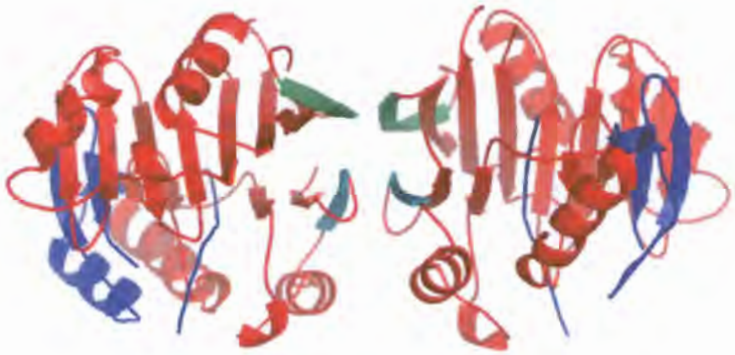

Figure 1.8: Dimer interface of human AdoMetDC. The α-subunits are coloured red, and the β-subunits blue. Green β-strand 15. Cyan β-strand 7. Generated in PYMOL. PDB ID: 1JEN.

naschii (Tolbert *et al.*, 2003a). Arginine decarboxylase and histidine decarboxylase do bear some structural similarities. In both structures the α- and β-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 β-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 β-strands from separate β-sheets. In AdoMetDC this occurs between two strands in the same β-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 β-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 it's 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.

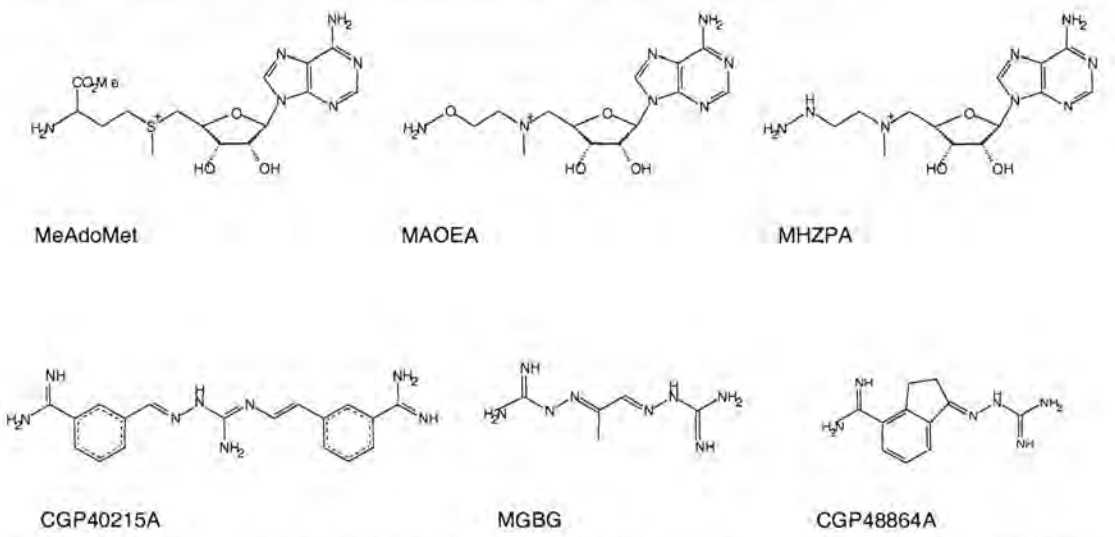

Figure 1.9: Known AdoMetDC inhibitors. MeAdoMet: S-adenosyl methyl ester, MAOEA: 5'-deoxy-5'-[*N*-methyl-*N*-[(2-aminooxy)ethyl]amino]adenosine, MHZPA: 5'-deoxy-5'-[*N*-methyl-*N*-(3-hydrazinopropyl)amino]adenosine, MGBG: Methylglyoxal bis(guanylhydrazone), CGP48864A: 4-amidinoindan-1-one-2'-amidinohydrazone.

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 α-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 β-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 it's effect via this series of charged residues. Glu11 may be required for the β-elimination step of processing, possibly mediating the deprotonation of the α-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 β-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 β-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 it's 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 (±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 by 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 it's 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.

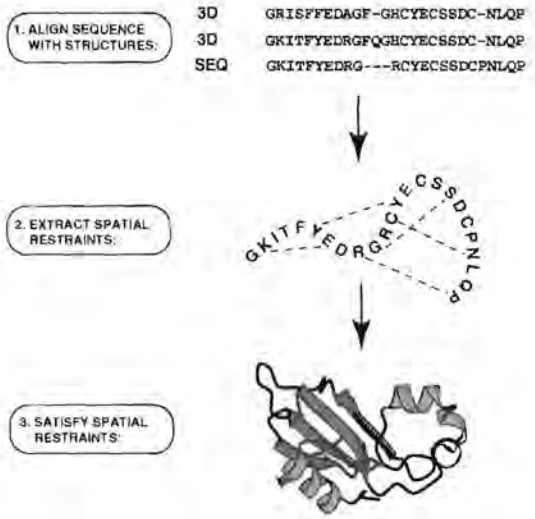

Figure 2.1: A brief overview of homology modelling (MODELLER manual). Firstly, the target sequence is aligned against the template sequences. This alignment is then used to devise the set of spatial restraints that the model of the target sequence must satisfy. The energy function representing the model 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 PLASMODB (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 PLASMODB 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 α-helices and β-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 α-chain, and the loop connecting helix 2 with β-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	http://www.compbio.dundee.ac.uk
PHD DoublePrediction HNNC SOPM SIMPA96 PREDATOR DSC Sec.Cons GOR4	http://npsa-pbil.ibcp.fr
SAM-T99	http://www.cse.ucsc.edu
HMMSTER	http://www.bioinfo.rpi.edu
PROF PROFSUB PHD PHDSUB PSIPRED	http://cubic.bioc.columbia.edu

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 $kcal.mol^{-1}.Å^{-1}$ was reached. This was followed in the second phase with conjugate gradients until a maximum derivative of 1 $kcal.mol^{-1}.Å^{-1}$ was reached. Results of modelling were analysed using LIGPLOT 4.1.1 (Wallace *et al.*, 1995), CERIUS2 (Accelrys ®) and PYMOL. The DSSP algorithm (Kabsch and Sander, 1983) was used to assign secondary structure as implemented in the dsspcmbi 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 PLASMODB with a smallest sum-probability (*P(N)*) of 1×10^{-219} , and from *P. yoelii*, chrPyl_cpy1465 was identified with a *P(N)* value of 4.9×10^{-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 ±4kb (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).

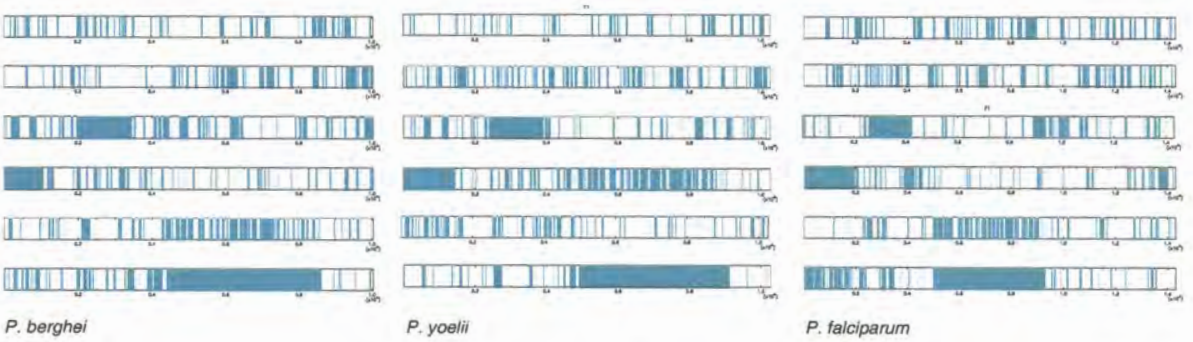

Figure 2.2: ORFs (turquoise) from the genomic regions identified from *P. berghei* and *P. yoelii*. Each row corresponds to another reading frame. All six reading frames are shown. Default parameters were used which resulted in all regions enclosed by start and stop codons. Each genomic region corresponds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

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).

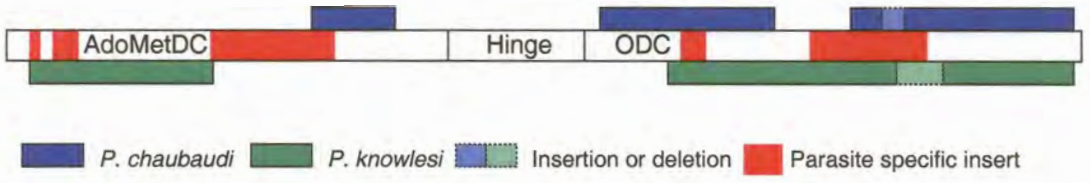

Figure 2.3: AdoMetDC fragments from *P. chabaudi* and *P. knowlesi*. The approximate location of fragments from *P. chabaudi* and *P. knowlesi* relative to the *P. falciparum* full bifunctional sequence is shown. *Plasmodium* specific inserts are also indicated.

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 (±43% identity between *P. falciparum* and the other sequences). The second insert is longer (±27 residues) and displays a higher degree of similarity (±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 (±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 (α-helices, 3-10-helices and β-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).

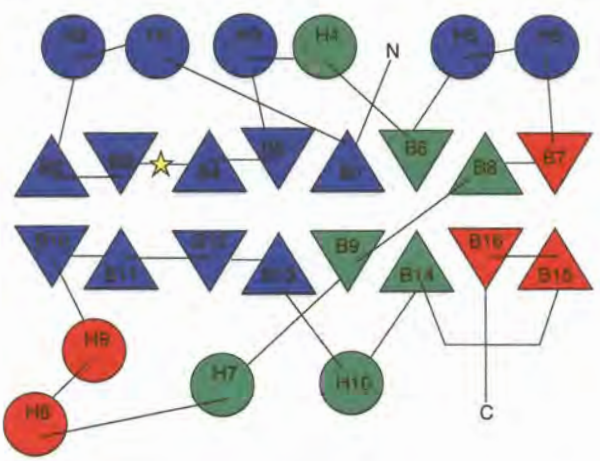

Figure 2.4: Homology of human secondary structural elements. The human topology diagram is used to indicate the degree of homology of human secondary structural elements and their predicted cognates in the malarial enzyme. Three alignments generated from `GONNET`, `PAM150` and `PAM250` scoring matrices were compared. Blue indicates 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 β-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 β-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 β-helix 6 and the 4 residues preceeding 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 β-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 α-helices 8 and 9 were manually adjusted in order to minimise the number of inserts thus creating one large insert of $\pm$150 residues. The majority of the insert was excised, leaving 2 residue overhangs. This insert is subsequently defined as insert 3 (Asp259-Phe408). Finally the regions corresponding to β-sheets 15 and 16 were manually adjusted in order to maximise the alignment of hydrophobic residues. The final alignment used for homology modelling (Fig. 2.5) still only retained modest sequence identity between the template sequences and the target sequences ($\pm$20%), despite removal of inserts 2 and 3.

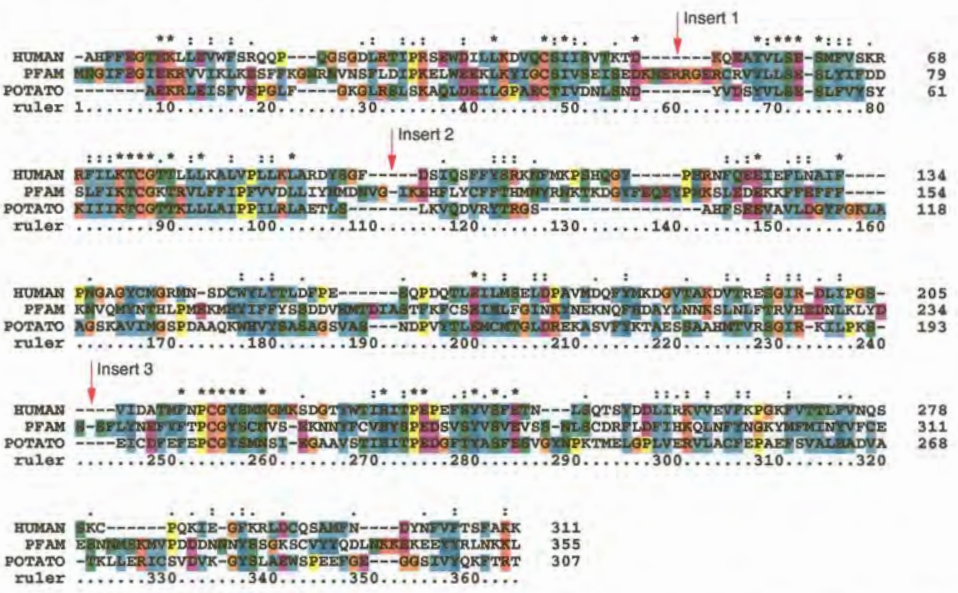

Figure 2.5: Final modelling alignment. The location of *Plasmodium* specific inserts is indicated by the red arrows. Standard `CLUSTALX` colouring is used (See App. A).

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 α-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 β-sheet and a C-terminal α-helix, with a coiled region in between (Fig. 2.6).

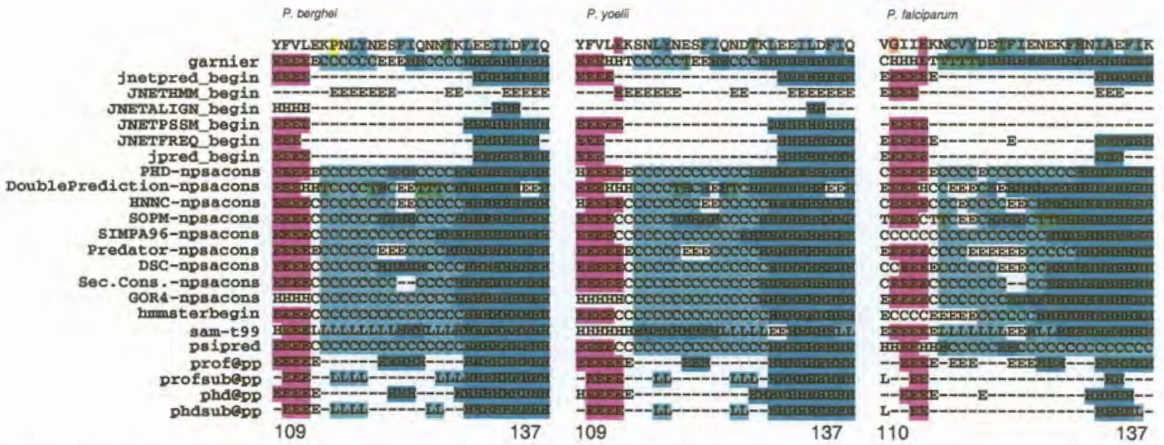

Figure 2.6: Predicted secondary structures of AdoMetDC insert 2. T: turn, C/L: coil, H: helix, E: β-sheet. Image generated with CLUSTALX using standard colours for protein sequences (See App. A).

The predictions for the long insert 3 are less similar. There appears to be a similar distribution and ordering of β-sheets and α-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 3.6×10^5 kcal.mol^{-1} (Final energy: -5664.8 kcal.mol^{-1}). 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
SDADKEVTTHIYSTRGTYEDTGMVNCVDVIYKNESTLLNRNNIENIPSIENKESNNNSRCCHNNNYSGSCHNIV
CCCCC-EEEEEEECCCCCCCCCEEEEEEEEECCCCC-CCCCCCCCCCCCCCCCCCCCCeeCCCCCCCCCCEEE

SVVPSERNNDHVHHRHYEDTLNRSNISAEDNNRNAQPEKEKDEDVRRDDEENKVLIKMIDTNLYECINYNKESF
EEE-CCCCCCCEee---CCCCCCCCCCCCCCCCCCCCCCCCC-HHhhhhHH--EEEEEECCCHeeEECCCCCC-
                                                                        408
```

P. berghei:

```
260
SEESEKIEVCNSSVYSFDSVENSSKYTSNTANTNITVNVSKDLDNNSINTLYNTIDDQTVMLTCSDDKNNNLNE
CCCCCEEEEECCCEEEECCCCCCCCCCCCCCCCCEEEEEEECCCCCCC-eeEEe-CCCCEEEEEECCCCCCCC-H

FDSKLLLAEENSKNMDKDNTQVFDNNNYIFISTPKA*QETYSNTNSVRSNEKSTCSSNTYTSLL*QNDLKEFHPKN
HHHHHHHHHH-CCCCCCCC-EEECCCCEEEEECCCCCC-CCCCCC-CCCCCCCCCCC------HHhHHH-CCCC

*NSMDTIEDDGK*LLVEEEVI*SYISNE*ETIDAKVEQLDNLSNKNMLDNNLDKNYEQTEANGPSLSSFTIACSDFLN
*CCCCC--CCC-*EEE*-----*EEeCC-*HHHHHHHHHHH-CCCC---CCCCCC----CCCCCCCC-*EEE*---HHHC

*KNANDEKGKDNTSSNMGIK*KIEENLYECINIQNN
*CCCCCCCCCCCCCCC----*HHHHHHH*----CCC
                                515
```

P. yoelii :

```
260
SEESEKIEVCSSVYSFDSVENSSKYTNNTANTNVTVNVSKDIDNNSINTLYNTVDDQTVILTCSDDKNNNLNE
CCCCCEEEEEEC-EEEECCCCCCCCCCCCCCC-EEEEEEECCCCCCC-eEEEE-CCC-EEEEEECCCCCCCCCC

SDSKLLLAGENSKNMDKDNTQVFNDDNNNNYIFISNPKV*QETYSNTNSVRSNEKSTCSSSSTY*TSLLQNDLKEF
C--Hhhhh-CCCCCCCCCC-EEECCCCCCCEEEE-CCCC-*EE-CCCCC-CCCCCCCCCCCCC-*HHHHHHHHHH-

*HPKNNAMDTIEDDGR*LLVEEEDI*SYISNE*ETIDAKVEQIDN*LSNKSILDNNLDKNYDQNEPSLSS*FTVAC*SDLI
*CCCCCCC----CCC-*EEEE*--C-*EEECC-*HHHHHHHHHHHCCCC--HCCCCCCCCCCCCCCCC-*EEEE*-*hhHh

*NKNANDEKSEDNTSKNIRIK*KIEENLYECINTQNN
*CCCCCCCCCCCCCCC-----*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: β-sheet, manual (by eye) assignment (i.e. weak consensus), h:α-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 β-sheet regions (red). The second set consists of three β-sheet regions (blue). Finally the third set of elements comprises an α-helical region followed by two β-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_α atoms between this initial model and templates were 1.45 and 1.65 Å for the human and potato templates, respectively (Table 2.2).

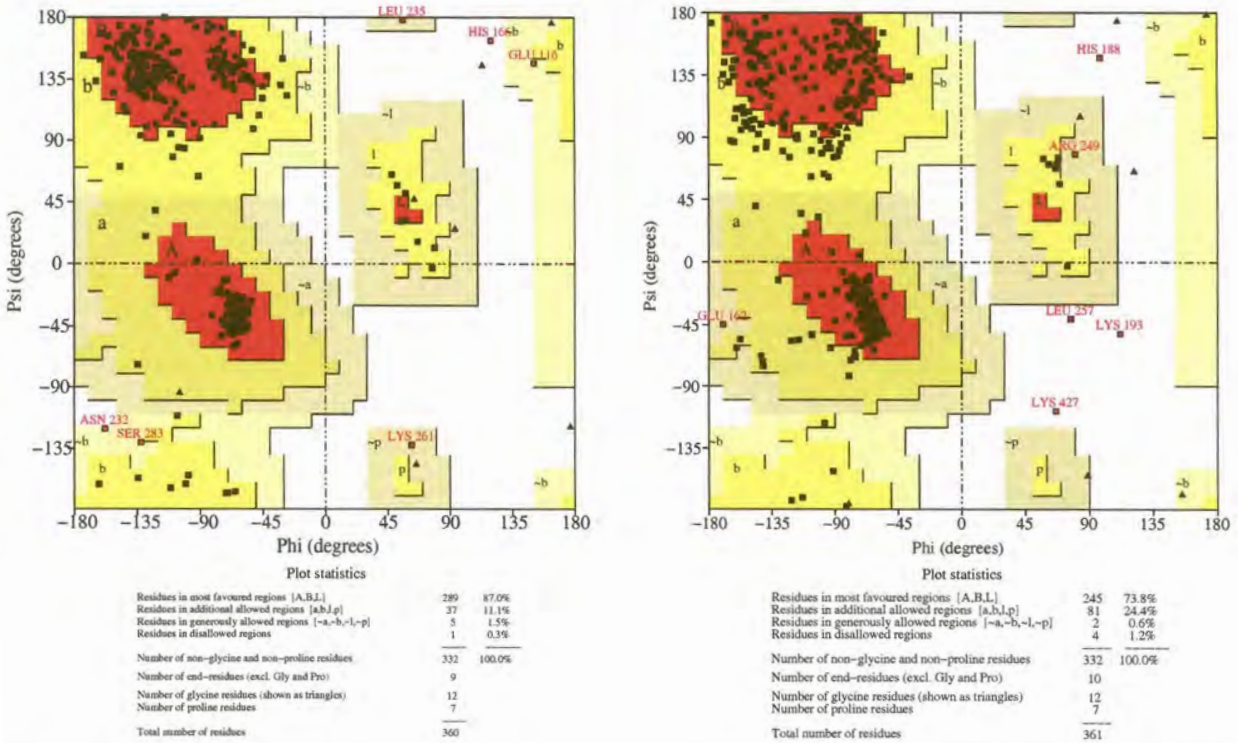

Figure 2.8: Ramachandran plots of the initial and final models. `MODELLER` model (left), final model after minimisation (right). Most sterically favoured regions (red), additional allowed regions (dark yellow), generously allowed regions (light yellow), disallowed regions (white). α-helix (A), β-sheet (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 β-sheets of 8 strands each, and 2 layers of helices. The helix layer lying adjacent to the β-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 α-helices showed most deviation, while the

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

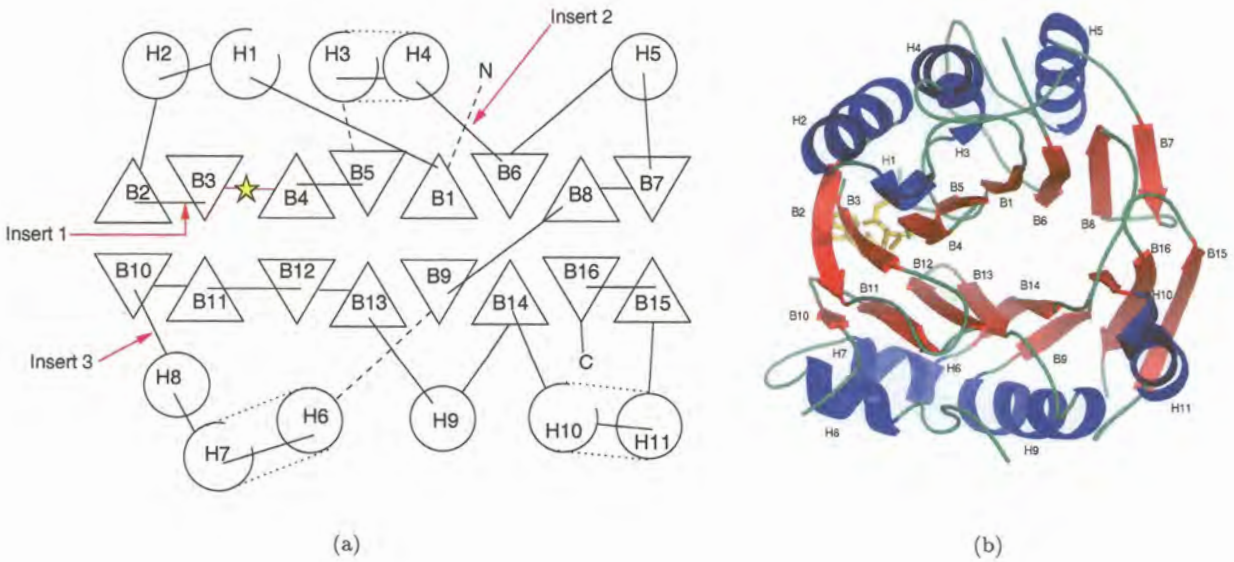

Figure 2.9: Topology of the final model. (a) Topology diagram, closed circles: α-helices, open circles denote: 3-10-helices, triangles: β-strands, dashed lines: 3-10-helix contiguous with an α-helix, star: site of proteolytic cleavage/active site. Inserts are as indicated. (b) Cartoon diagram, β-strands: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

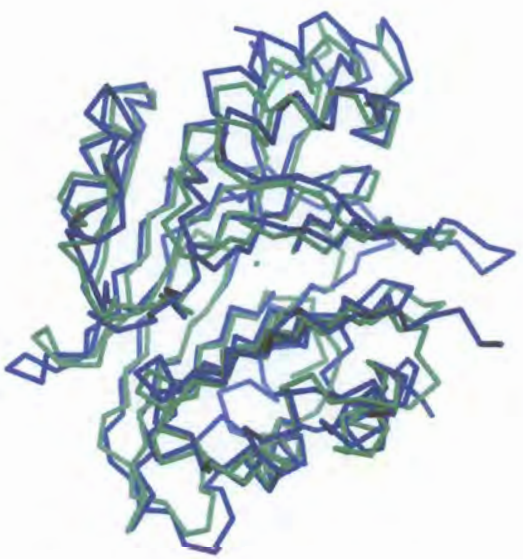

Figure 2.10: C-α trace superimposition of model (blue) on human template (green).

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*hum* is a

catalytically important residue in the human enzyme, most likely required for protonation of the α-carbon of the substrate cysteinyl moiety after decarboxylation (Xiong *et al.*, 1999). The sulphydryl 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.

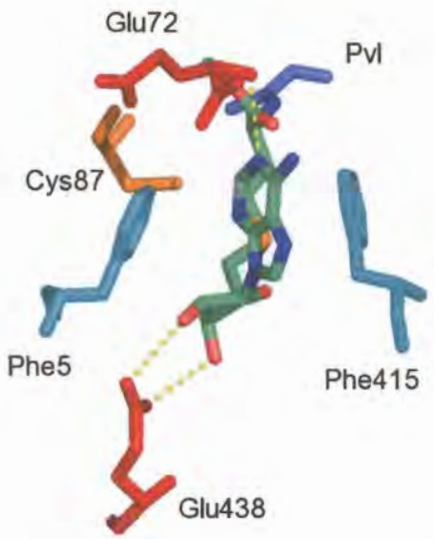

Figure 2.11: Important model-active site residues. Residues are coloured according to their overall nature: acidic residues (red), aromatic (cyan), polar (orange), modified (slate). Pvl: pyruvoyl. The ligand (MeAdoMet) is coloured by atom: N (blue), O (red), C (green), S (orange). Hydrogen bonds are 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 2Å 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
(*Gly3*)	(*His5*)	Ser74	Ser69	Tyr420	Tyr228
Phe5	Phe7	(Thr86)	Thr81	Ser421	Ser229
(Glu6)	(Glu8)	Cys87	Cys82	(His434)	(His243)
(Glu9)	(Glu11)	Thr90	Thr85	*Tyr435*	*Ile244*
Leu70	Leu65	Phe415	Phe223	Pro437	Pro246
Ser71	Ser66	*Thr416*	*Asn224*	Glu438	Glu247
Glu72	Glu67	Cys418	Cys226	(*Ser436*)	*Thr245*
Pvl73	Pvl68	Gly419	Gly227		

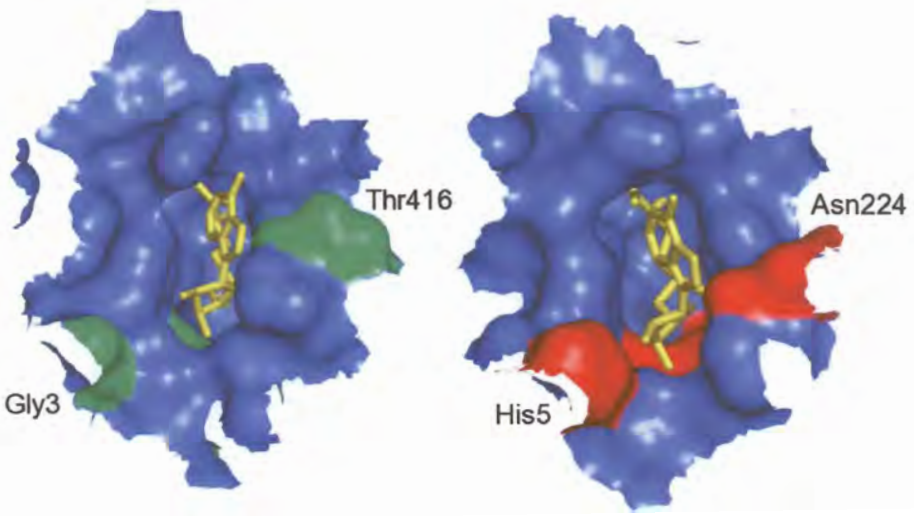

Figure 2.12: Substitutions in the model active site. Van der Waals surfaces of the model (left) and human (right) active sites are shown. Substituted residues are indicated in green and red. Tyr435, Ser445, Ile224*hum* and Thr245*hum* (unlabelled) are visible behind the MeAdoMet ligands (yellow). Surfaces 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 β-strand 11, while in the model it points between Tyr435 of β-strand 13.

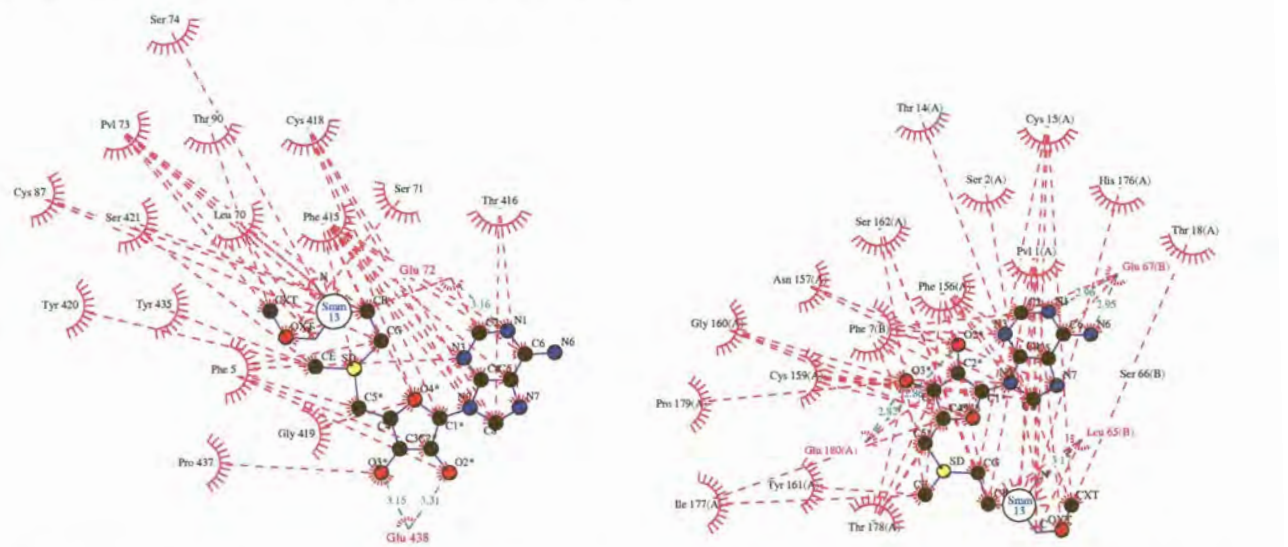

Figure 2.13: Ligand interactions for the model (left) and human (right) active sites with MeAdoMet. "Hedgehogs" indicate hydrophobic interactions. Hydrogen bonds are indicated by green dashed lines.

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*hum* 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*hum*, and is surrounded by Gly9*hum*, Glu11*hum*, Ser254*hum*, His243*hum*, Ile244*hum* and Thr245*hum*. 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 glutamate residue (Glu11*hum*, Glu9). The glutamate residue is conserved across all species, and situated between 8 and 9 Å from the sulphonium atom.

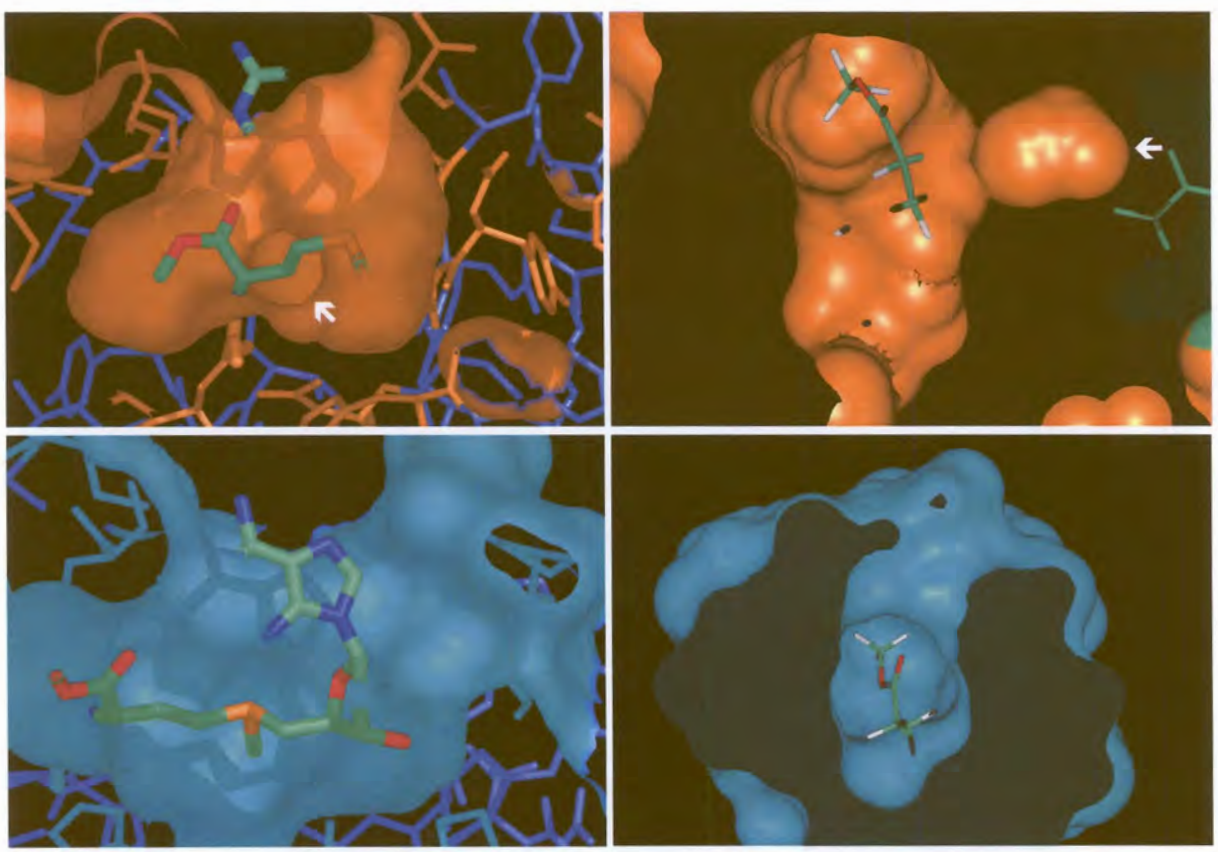

Figure 2.14: Active site shapes of AdoMetDC. Top: model active site. Bottom: human active site. The MeAdoMet ligand is included: C (green), N (blue), S (orange), O (red). Surfaces generated in INSGIHTII (right) and PYMOL (left). White arrows indicate extra cavities.

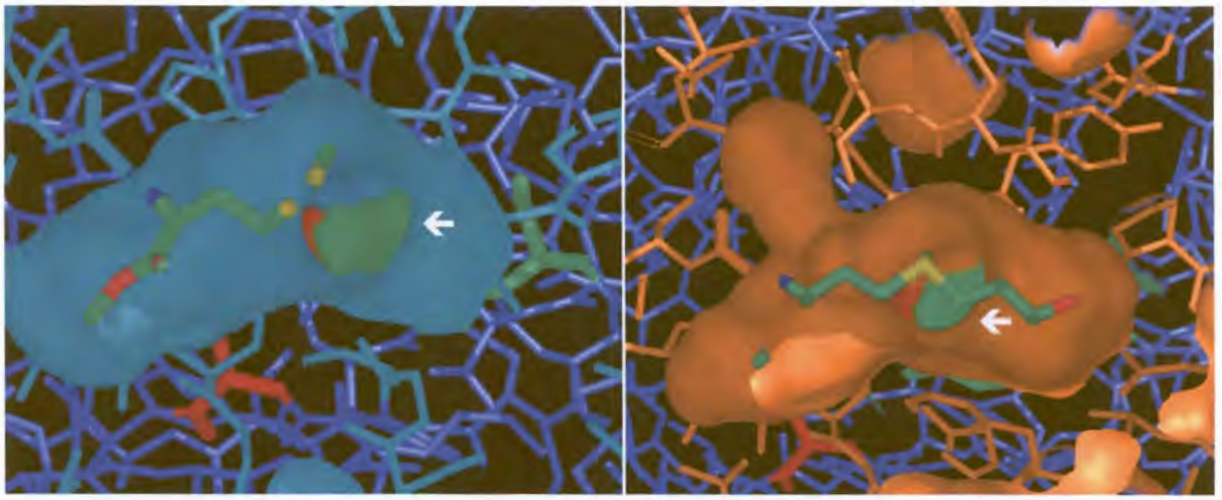

Figure 2.15: Cavity identified near the sulphonium group (white arrow). Human (left) and model (right) structures. C (green), S (yellow), N (blue), O (red). Surfaces generated with PYMOL.

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*hum* and Arg33*hum* from helix 1 are within electrostatic bonding distance of each other. Arg33*hum* is also situated close to Asp58*hum* and Asp60*hum*. Furthermore, Glu61*hum* is located adjacent to Lys56*hum* in β-strand 2. In the potato structure this network is not as extensive, with Arg38*pot* (equivalent of Arg33*hum*) being located near to Asp63*pot* (Asp58*hum*) and Asp66*pot* (Glu61*hum*). The corresponding potato residues for Asp31*hum* and Lys56*hum* are not similarly charged (Gly and Tyr, respectively). A more extensive charge network is seen in the model. Arg38*pot*/Arg33*hum* is replaced by Leu31. Glu61*hum*/Asp66*pot* and Asp58*hum*/Asp63*pot* 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*hum*) may also possibly interact with Arg66, but this residue is possibly too far away ($\pm$ 11 Å), and it's 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*.

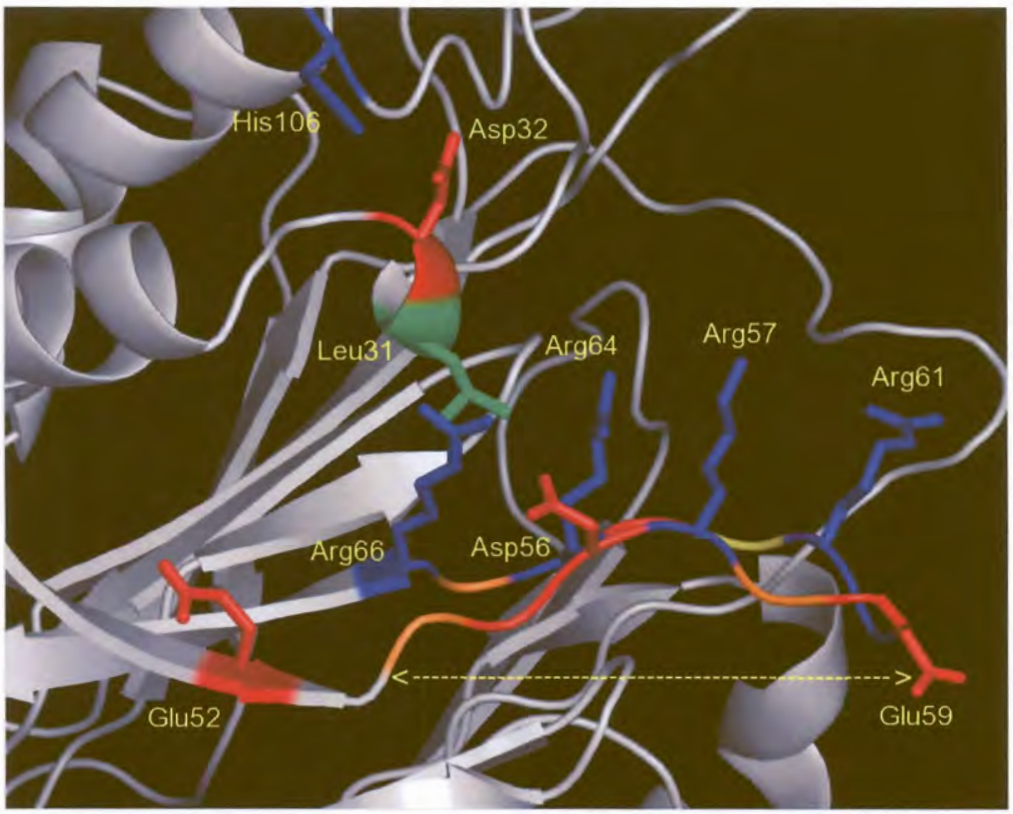

Figure 2.16: Charge network associated with insert 1. Acidic (red), basic (blue), hydrophobic (green) residues referred to in section 2.3.4.4 are displayed as stick models. Insert 1 is indicated by the arrows.

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 (±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 (± 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 (β-strands 7, 15 and 16), occupy the region responsible for dimerisation within the human enzyme. Furthermore, β-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 β-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 β-strand in the N-terminal region and an α-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. α-helix : β-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 it's 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 it's 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 it's 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 (±20 % after removal of inserts 2 and 3). The RMSD of the C-α 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 ±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 ϕ and ψ 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 (±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. β-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 β-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 α-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 α-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 sulphydryl group of Cys87 points towards the bound inhibitor, while that of Cys87*pot* is orientated away. Secondly, Glu72*pot* is more embedded in the active site pocket, while Glu72 and Glu67*hum* 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*pot* 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*hum* is replaced by the less bulky Ser436 and may therefore explain why this residue does not contact the ligand. Adjacent to this position, Ile242*hum* 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*hum*. The last observable substitution, which is not predicted to make contact in either humans or the model, is that of Gly3 replacing His5*hum*. 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*hum* 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^1 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^+ 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 it's 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 it's 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.