

## Chapter 3

# Model guided mutational analysis of malarial

## AdoMetDC

### 3.1. Introduction

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

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

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

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

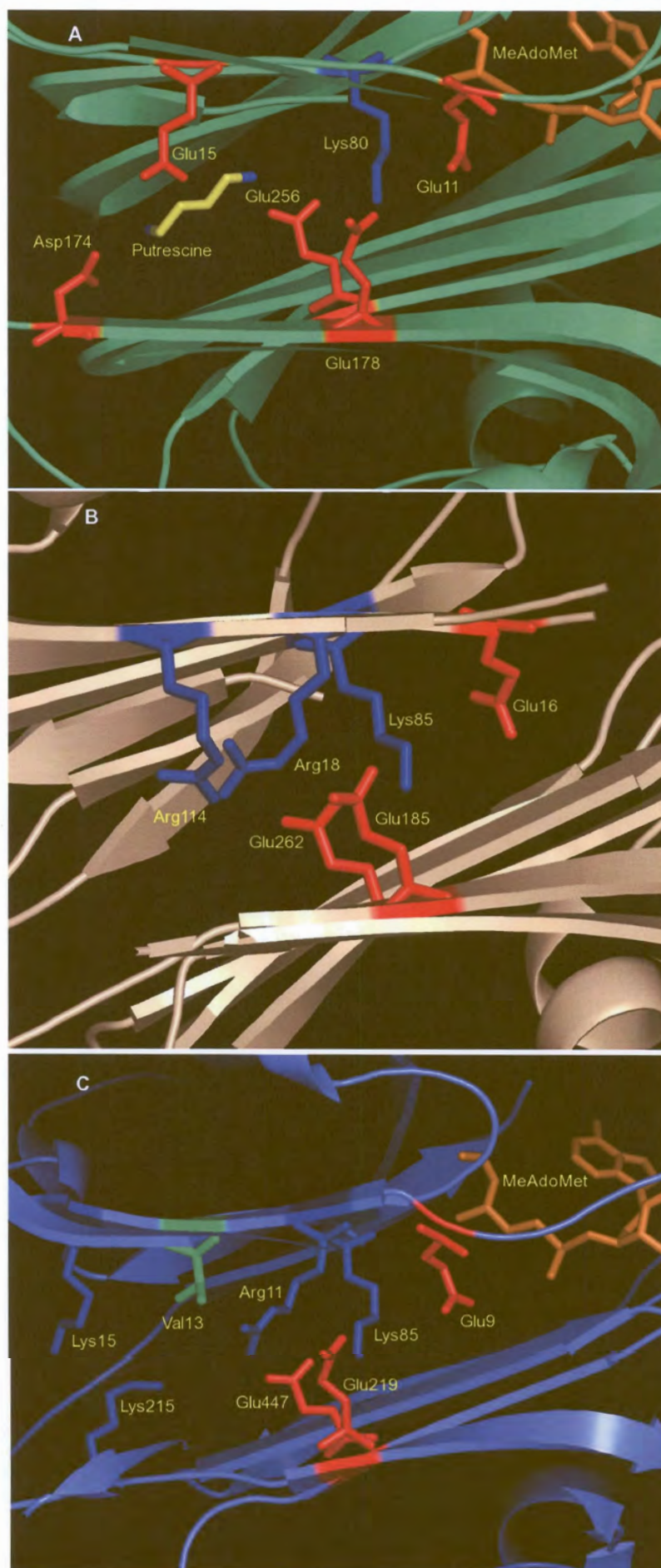


Figure 3.1: Putrescine charge networks. (A) Human (PDB ID: 1I7B). (B) Potato (PDB ID: 1MHM). (C) *P. falciparum* model. Associated acidic residues (red), basic residues (blue) and hydrophobic residues (green) are indicated. MeAdoMet (orange) and putrescine (yellow, blue) are also shown.



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

## 3.2. Methods

### 3.2.1. *In silico* putrescine docking

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

### 3.2.2. Construction of putrescine-like mutants

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

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

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

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

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

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

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

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

### 3.2.2.2. Sequencing of putrescine-like mutants

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



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

### 3.2.3. Recombinant expression

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

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

### 3.2.4. Enzyme Assays

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

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

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

## 3.3. Results

### 3.3.1. Putrescine docking

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

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



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

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

### 3.3.1.2. Putrescine docking

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

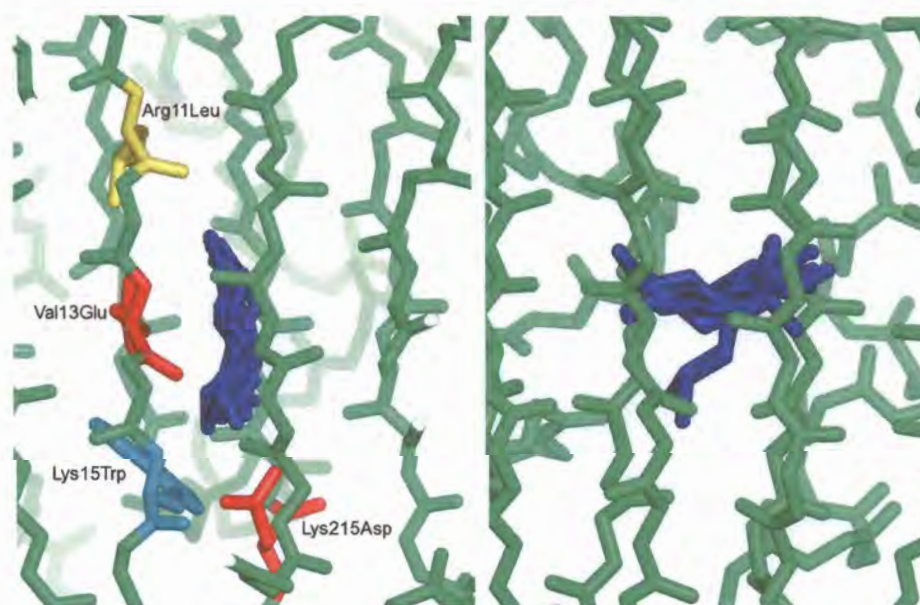


Figure 3.2: Orientation of putrescine. Chimeric (left) and wild-type (right) models. The ten best (lowest energy) conformations of putrescine (blue) are shown for each. *In silico* mutations for the chimeric model are separately coloured.

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

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

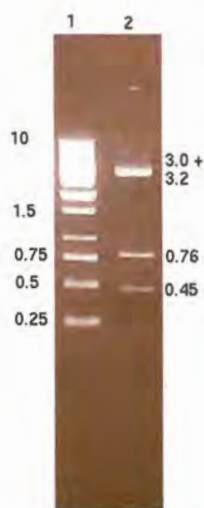


Figure 3.3: *Xba*I and *Hind*III restriction of AdoMetDC/ODC cloned into pASK-IBA3 (right). Lane 1: 1 kb ladder (Fermentas ®). Lane 2: digestion of expression plasmid. Sizes are given in kbp.

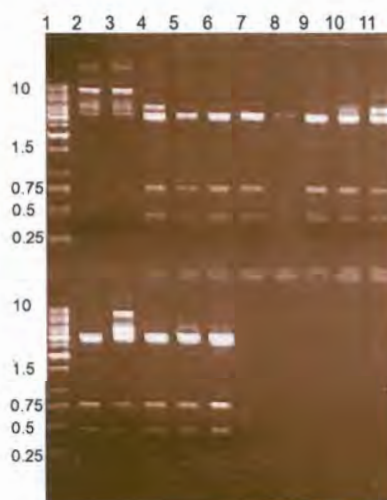


Figure 3.4: *Xba*I and *Hind*III restriction of Arg11Leu (lanes 2-6 top), Lys15Ala (lanes 7-11 top) and Lys215Ala (lanes 2-6 bottom). Marker lanes = lane 1 top and bottom. Incomplete digestion occurred in the first two lanes of Arg11Leu. Sizes are given in kbp.



### 3.3.3. Expression of mutant AdoMetDC/ODC

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

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

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

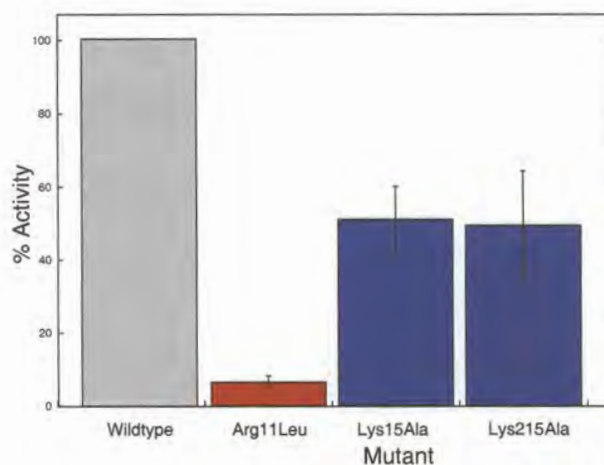


Figure 3.5: Effect of mutations on AdoMetDC activity. Mean relative activities and standard deviations for each mutant are derived from all observations used to construct Table 3.3. Relative activity was calculated from specific activity for each mutant relative to the wild-type activity from that particular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

## 3.4. Discussion

### 3.4.1. Putrescine docking

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

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

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

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



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

### 3.4.2. Mutagenesis

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

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



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

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



## Chapter 4

# Model guided inhibitor screening of malarial

## AdoMetDC

### 4.1. Introduction

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

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

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

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

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

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

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



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

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

## 4.2. Methods

### 4.2.1. *In silico* inhibitor screening

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

### 4.2.2. Test compound solutions

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

### 4.2.3. Assays

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

## 4.3. Results

### 4.3.1. *In silico* inhibitor screening

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

### 4.3.2. Solubility of potential inhibitors

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



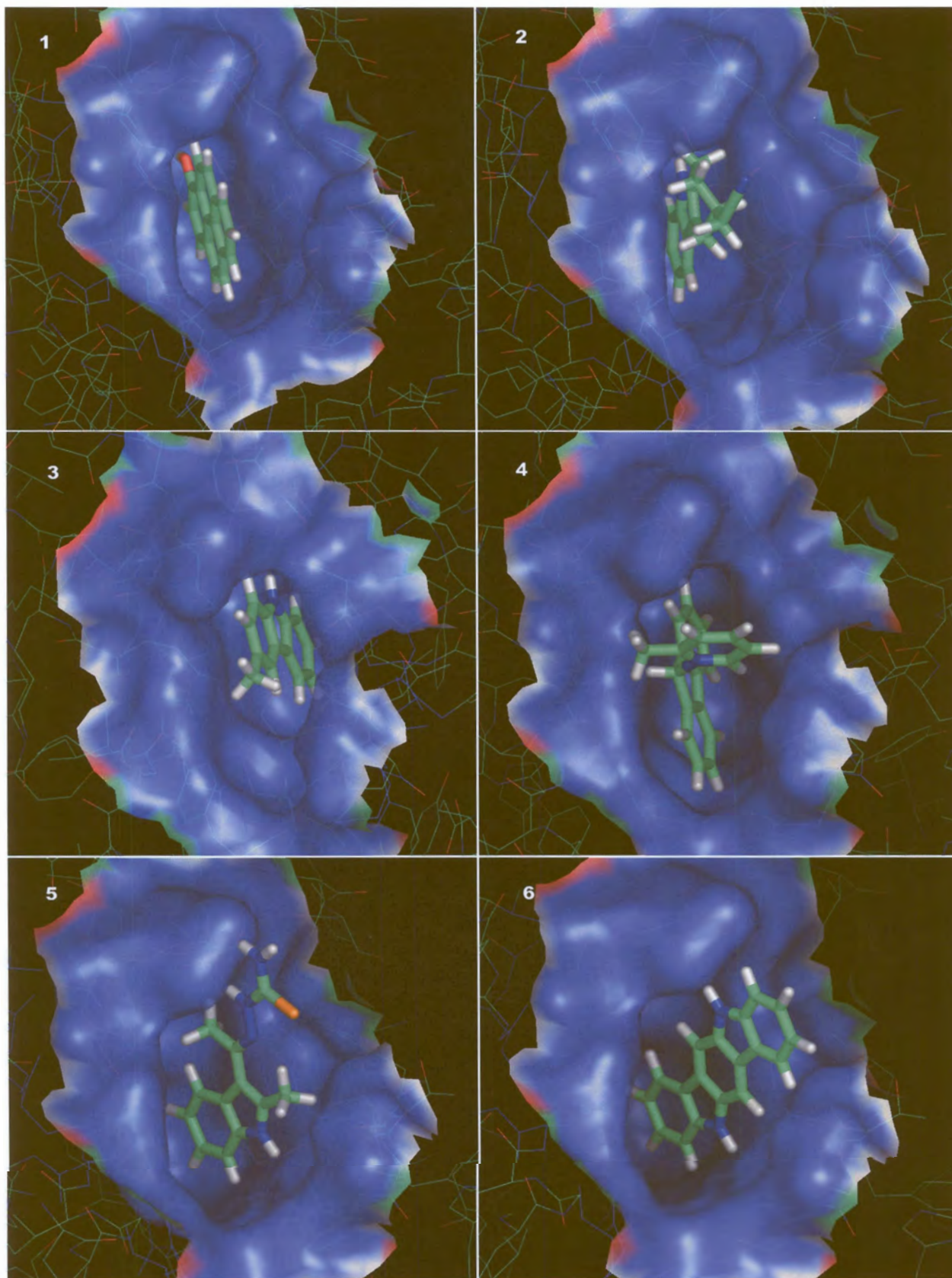

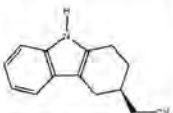
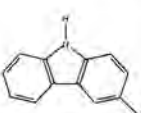
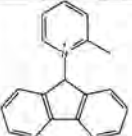
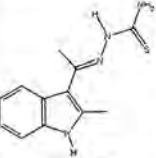
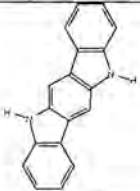
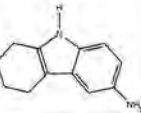
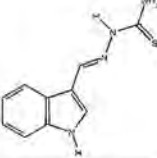
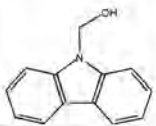
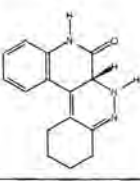
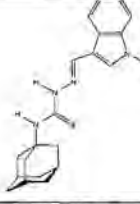


Figure 4.1: Orientations of the top 6 potential NCI inhibitors.

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

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

#### 4.3.3. Inhibition of AdoMetDC

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

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

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



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

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

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

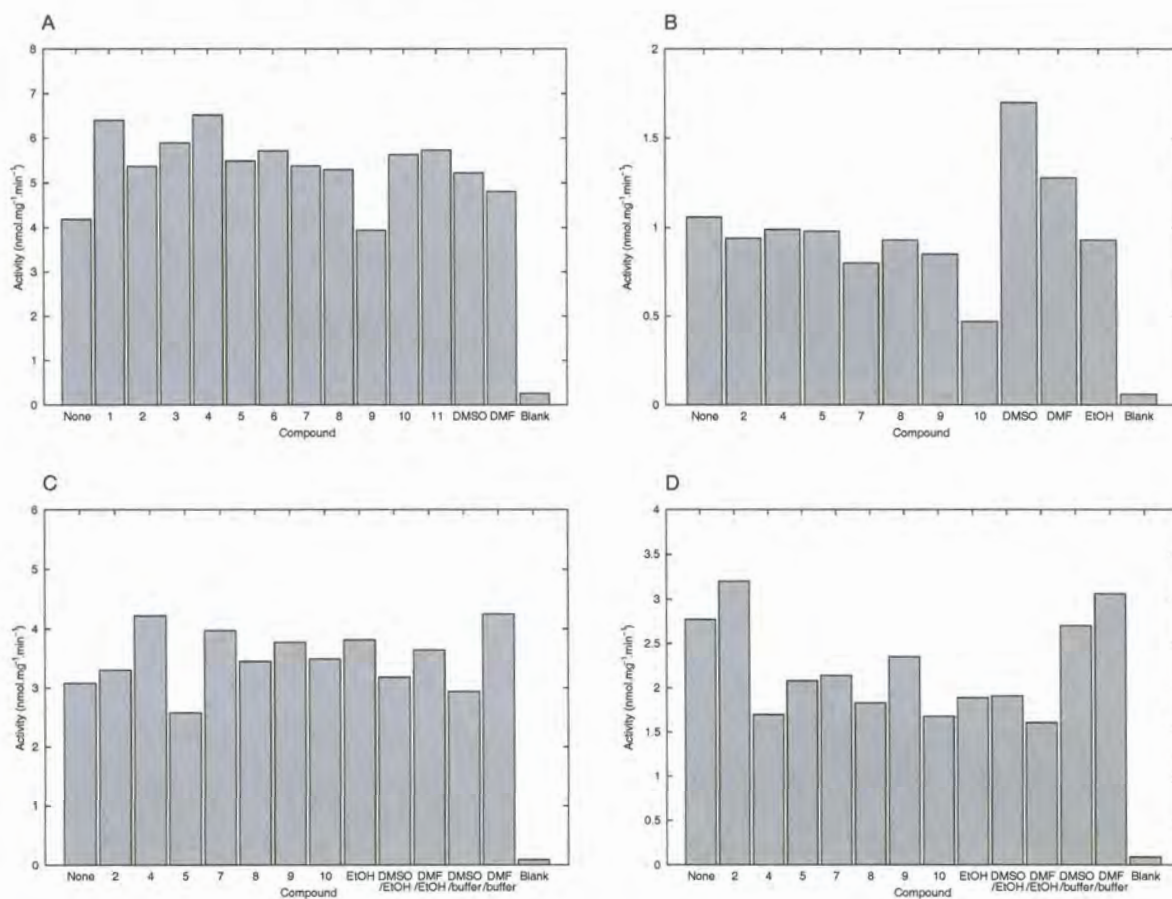


Figure 4.2: Overview of effect of identified inhibitors and solvent controls on AdoMetDC activity. (A) Equimolar test compound and substrate, excess solvent controls. (B) Equimolar test compound and substrate, excess solvent controls, 1/10 substrate/test-compound concentration. (C) Equimolar test compound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

#### 4.4. Discussion

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

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



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

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

## Chapter 5

# Concluding Discussion

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

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

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



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

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

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

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

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



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

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

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



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

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

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

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

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