CHAPTER 4

Functional and structural roles of parasite-specific inserts in the bifunctional S-adenosylmethionine decarboxylase/ornithine decarboxylase.

4.1) INTRODUCTION.

Numerous *P. falciparum* proteins are characterised by an increased protein size relative to homologues from other organisms (Bowman, *et al.*, 1999; Gardner, *et al.*, 1998). Factors contributing towards this increased protein size include the presence of unique parasite-specific regions that intersperse conserved areas of other protein homologues as well as the peculiar bifunctional organisation of some malarial proteins, with two protein activities residing on a single polypeptide. The PfAdoMetDC/ODC presents with both of these characteristics (Chapter 3).

Different sized inserts are found in malarial protein kinases (Kappes, *et al.*, 1999), HSP 90 (Bonnefoy, *et al.*, 1994), RNA polymerases (Giesecke, *et al.*, 1991), dihydrofolate reductase-thymidylate synthase (DHFR-TS) (Bzik, *et al.*, 1987), glutathione reductase (Gilberger, *et al.*, 2000), γ-glutamylcysteine synthetase (Luersen, *et al.*, 1999) and the P-Type ATPase 3 (Rozmajzl, *et al.*, 2001). The precise function and evolutionary advantage of these inserts remain unclear. Some speculations for the functions of these inserts include possible interaction sites with as yet undefined regulatory proteins in the parasite, interaction sites with host proteins and a method to evade the host immune response (Li and Baker, 1998; Schofield, 1991). Speculations on the presence of inserts in *P. falciparum* proteins include simple evolutionary divergence that may not necessarily affect the activity and/or structure of the protein. However, strong selective pressures must exist to maintain and diversify these regions (Ramasamy, 1991).

The parasite-specific areas are normally characterised by repetitive, highly charged amino acid stretches (Chapter 3, (Pizzi and Frontali, 2001). In particular, Asn- and Asp-rich areas have been characterized in antigenic regions of membrane proteins and are speculated to play a role in evasion of the host defence mechanisms by acting as

antigenic smokescreens (Barale, *et al.*, 1997; Kemp, *et al.*, 1987; Reeder and Brown, 1996). *P. falciparum* has various Asn-rich proteins in particular STARP (sporozoite threonine and asparagine rich protein) (Facer and Tanner, 1997), the clustered asparagine rich protein (CARP) (Wahlgren, *et al.*, 1991) and the circumsporozoite protein (PNANP repeat) (Kwiatkowski and Marsh, 1997). These repeats are present in immunodominant domains associated with antigenic proteins on the surface of the parasite. The parasite also has other proteins rich in specific amino acids including the histidine-rich protein (Kwiatkowski and Marsh, 1997), the glutamate-rich protein (Hogh, *et al.*, 1993), a histidine-alanine rich protein (Stahl, *et al.*, 1985) and a serine repeat protein (Kwiatkowski and Marsh, 1997). These sequences probably all relate in some way to the structure-activity properties of these proteins.

Bifunctional proteins are not unusual in *P. falciparum* and indeed also in other parasitic protozoa. The malaria parasite has several bifunctional enzymes including DHFR-TS (also found in *L. donovani* and *T. brucei*)(Bzik, *et al.*, 1987; Ivanetich and Santi, 1990), dihydropteroate synthetase-dihydrohydroxymethylpterin pyrophosphokinase (DHPS-PPPK)(Triglia and Cowman, 1994), glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase (Clarke, *et al.*, 2001) and guanylyl cyclase-adenylyl cyclase (Carucci, *et al.*, 2000). Various speculations have been put forward to explain the bifunctional nature of these proteins. In the case of DHFR-TS, the two proteins catalyse consecutive reactions in the same metabolic pathway and substrate channelling has been proposed to optimise formation of products without further regulatory processes involved (Ivanetich and Santi, 1990). Other possible explanations for the bifunctional arrangements include coordinated regulation of protein concentrations/activities and intramolecular communication and interaction (Müller, *et al.*, 2000).

Obvious questions arise as to the importance of the parasite-specific inserts in the activity and/or structure of the bifunctional PfAdoMetDC/ODC. Much is known about certain key residues in PfAdoMetDC/ODC from mutagenesis results (Krause, *et al.*, 2000; Müller, *et al.*, 2000; Wrenger, *et al.*, 2001). Point mutations in one domain of the complex do not influence the activity of the other. Similarly, inhibition of one domain with a specific inhibitor has a singular effect on that domain. It therefore seems that the individual decarboxylase activities can function independently from each other (Wrenger, *et al.*, 2001). However, certain protein-protein interactions are expected in order to stabilize the bifunctional complex. One possible role for the inserted amino

acids and/or hinge region could be to mediate these protein-protein interactions. Clarification of the possible functions of these parasite-specific areas could contribute towards understanding the properties of PfAdoMetDC/ODC and exploitation of this knowledge in the design of selective inhibitors for antimalarial chemotherapy.

One of the most powerful developments in molecular biology has been the ability to create defined mutations in a gene and to analyse these effects on the activities of *in vitro* expressed mutated proteins. Mutants are essential in understanding the structure-function relationships of proteins and aid the rational design of proteins and their inhibitors (Lodish, *et al.*, 1995). Numerous methods are available for site-directed *in vitro* mutagenesis of genes, collectively termed protein engineering (Old and Primrose, 1994; Wilson and Walker, 2000). Mutations are designed to alter a particular codon or stretch of codons, which after translation give rise to different amino acids that may influence the properties of the protein. Cassette-mutagenesis results in the replacement of a particular DNA sequence with a synthetic DNA fragment containing the desired mutation with almost 100% efficiency (Old and Primrose, 1994). The disadvantage of this technique is the requirement of unique restriction sites flanking the region of interest. Oligonucleotide or single-primer mutagenesis (primer-extension) requires single stranded DNA (insert cloned in e.g. M13 phage vectors) from which DNA synthesis is primed with the mutant oligonucleotide. Subsequent cloning of the products produces multiple copies, half of which are mutants and half wild type, necessitating large-scale screening procedures (Old and Primrose, 1994). PCR mutagenesis relies on the incorporation of mismatches in the PCR primers into the amplified product (Wilson and Walker, 2000). PCR mutagenic methods include techniques termed overlap-extension PCR (where two primary PCRs produce two overlapping fragments that are subsequently amplified) and megaprimer PCR mutagenesis (where the products of the primary PCR are allowed to act as primers in the subsequent amplifications). Various modifications of these methods allow rapid mutagenesis at almost 100% efficiency (Old and Primrose, 1994).

This chapter describes results of studies aimed at elucidation of the role of parasite-specific inserts in interactions between the AdoMetDC and ODC domains in the bifunctional PfAdoMetDC/ODC enzyme. Our strategy was to utilise site-directed *in vitro* mutagenesis methods to gauge their effects on complex formation and activities of the two domains.

Some of the results obtained in this Chapter have been submitted for publication in the Biochemical Journal (Birkholtz, *et al.*, 2002b).

4.2) MATERIALS AND METHODS.

4.2.1) Amino acid sequence and structural analyses.

Amino acid sequence alignments were performed with Clustal W (Thompson, *et al.*, 1994) using the default parameters for PfAdoMetDC/ODC (Genbank Accession Number AF094833) and the corresponding enzymes from the human, mouse, *L. donovani* and *T. brucei*. Genbank accession numbers for AdoMetDCs: human: M21154, murine: D12780, *T. brucei:* U20092, *L. donovani:* LDU20091 and for ODCs: human: M31061, murine: J03733, *T. brucei:* J02771 and *L. donovani:* M81192. Bifunctional AdoMetDC/ODC was also identified in other *Plasmodium* species as described in Chapter 3 and these sequences were also included in the multiple alignment. Secondary structure predictions, antigenic profiles and Kyte and Doolittle hydrophobicity plots were obtained with the PredictProtein server (Rost, 1996).

4.2.2) Deletion mutagenesis (Kunkel, 1985).

Deletion mutants were created for all the major inserts present in both the PfAdoMetDC and PfODC domains, as well as for the hinge region connecting these domains in the bifunctional enzyme. Mutagenesis was based on the principle described in the QuikChange™ Site-Directed Mutagenesis Kit by Stratagene (La Jolla, California, USA). Briefly, PCR is used to introduce site-specific mutations to any double-stranded supercoiled plasmid containing the insert of interest. Two complementary mega-primers with the desired mutations are used to create mutated plasmids with staggered nicks after linear amplification. *Pfu* DNA polymerase from *Pyrococcus furiosis* is used to replicate both plasmid strands with high fidelity using its 3'-5' proofreading exonuclease activity without displacing the mutant primers. The product is then treated with *Dpn*I (target sequence: 5'-Gm^{6}ATC-3') in order to remove the methylated parental DNA template. The PCR-generated mutated plasmid is then transformed into competent *E. coli* cells where the bacterial ligase system repairs the nicks to create double stranded plasmids. This technique combines the principle of oligonucleotide mutagenesis with

PCR-based techniques to obtain a >80% efficiency in mutagenesis of any insert in any vector system.

Oligonucleotides used for the site-directed deletion mutagenesis are indicated in Table 4.1. A typical deletion mutagenesis reaction (50 µl final volume) contained 10 ng of the wild-type expression plasmids (as isolated in Chapter 3, section 3.2.2) with the specific inserts (pASK-IBA3 for bifunctional PfAdoMetDC/ODC; pASK-IBA7 with either the PfAdoMetDC or PfODC domains; Institut für Bioanalytik, Göttingen, Germany), 150 ng of both the mutagenic sense and antisense mega-primers (Table 4.1), 1x *Pfu* DNA polymerase reaction buffer, 2.5 mM of each dNTP and 3 U *Pfu* DNA polymerase (Promega, Wisconsin, USA). The cycling parameters were 95°C for 50 sec, 55°C for 1 min and 68°C for 12 min (bifunctional construct) or 9 min (separate domains) repeated for 18 cycles in total after an initial denaturation step of 95°C for 3 min in a Perkin Elmer GeneAmp PCR system 9700 (PE Applied Biosystems, California, USA). After 9 cycles, a further 1 U *Pfu* DNA polymerase was added to amplifications of the bifunctional constructs. The PCR products were subsequently treated with 20 U *Dpn*I (New England Biolabs, Massachusetts, USA) for 3 hours at 37°C followed by removal of the digested parental DNA templates using the standard protocols described in the High Pure PCR Product Purification Kit (Roche, Mannheim, Germany). The pure mutated constructs containing nicks were ligated at 4°C for 16 hours with 6 U T4 DNA ligase (Promega, Wisconsin, USA) to increase the transformation efficiency. The double stranded supercoiled plasmids were subsequently transformed into competent DH5α *E. coli* cells (Invitrogen, Paisley, UK) as described in Chapter 2, section 2.2.10.

4.2.3) Nucleotide sequencing of the various mutants.

The nucleotide sequences of the mutant cloned fragments were determined by automated nucleotide sequencing as described in Chapter 2, section 2.2.11. For cycle sequencing of mutant clones, primers complementary to the PfAdoMetDC/ODC nucleotide sequence was used at a site not more than 300 nucleotides removed from the mutation site. See Chapter 2, Fig. 2.3 for primer locations.

4.2.4) Recombinant expression and purification of wild type and mutant proteins.

The *P. falciparum* monofunctional AdoMetDC and ODC and bifunctional AdoMetDC/ODC were expressed as Strep-Tag fusion proteins as described in Chapter 3, section 3.2.2 (Krause, *et al.*, 2000; Müller, *et al.*, 2000; Wrenger, *et al.*, 2001). Mutant forms of PfAdoMetDC/ODC with individual deletion of the parasite-specific inserts, as well as single and combined insert deletion mutants in the monofunctional PfAdoMetDC and PfODC domains were isolated as for the wild type proteins. The monofunctional PfODC domain lacking the N-terminal hinge region that connects it to PfAdoMetDC was cloned into the expression plasmid pJC40 for expression as a fusion protein with an N-terminal His_6-tag (Krause, *et al.*, 2000). The entire coding region of the *P. falciparum* spermidine synthase was also cloned in the same plasmid (Haider *et al*, personal communication). These proteins were isolated as described in Krause *et al.* 2000. The concentrations of the purified proteins were determined with Coomassie brilliant blue G-250 (Pierce, Illinois, USA) as described by Bradford (Bradford, 1976). Purified proteins were analysed by SDS-PAGE and visualised with silver staining as described in Chapter 3, section 3.2.6.

4.2.5) Protein-protein interaction determinations.

4.2.5.1) Size-exclusion fast protein liquid chromatography (SE-FPLC) of the interacted proteins.

Protein-protein interactions were determined by SE-FPLC of wild type or mutant hinge-linked bifunctional proteins and combinations of individually expressed wild type or mutant monofunctional PfAdoMetDC and PfODC. Wild type and mutant forms of PfAdoMetDC/ODC were analysed for their ability to form heterotetrameric complexes (~330 kDa) or uncomplexed heterodimer subunits (~160 kDa). Combinations of wild type and mutant monofunctional PfAdoMetDC and PfODC were analysed for their ability to associate and form hybrid heterotetrameric complexes (~330 kDa), to remain in their monofunctional active states (heterotetrameric PfAdoMetDC of ~ 145 kDa and homodimeric PfODC of ~166 kDa) or heterodimeric ~ 64 kDa AdoMetDC and ~80 kDa monomeric ODC. Intermolecular protein-protein interactions between bifunctional PfAdoMetDC/ODC and spermidine synthase were also analysed. Separately expressed and isolated proteins were allowed to interact by co-incubation for 10 min at room temperature. Subsequently, the protein complexes were subjected to FPLC as described in Chapter 3, section 3.2.4. Protein was detected in the collected fractions with

Coomassie brilliant blue G-250 (Pierce, Illinois, USA), dot-blot western immunodetection or enzyme activity determinations as described in the next sections.

4.2.5.2) Dot-blot Western analyses of SE-FPLC fractions.

The collected size-exclusion chromatography fractions were transferred to nitrocellulose membranes using a BioDot apparatus (Bio-Rad) and analysed by dot blot Western. The membranes were blocked in 3% w/v low fat milk powder in 1xPBS for 16 hours at 4°C followed by incubation with a 1:4000 dilution of polyclonal Strep-tag II rabbit antiserum raised against the Strep-tag II peptide conjugated to keyhole limpet hemocyanin (Institut für Bioanalytik, Göttingen, Germany) for 1 hour at room temperature. After three washes with 0.05% Tween-20 (Merck, Germany) in 1xPBS, the membrane was incubated with a 1:2000 dilution of horseradish peroxidase (HRP) conjugated anti-rabbit donkey whole IgG (Amersham Pharmacia Biotech, UK) for 1 hour at room temperature in 1% w/v low fat milk powder in 1xPBS. The membrane was again washed three times in 0.05% Tween-20 in 1xPBS. The proteins were visualized with the ECL Plus™ Western Blotting system (Amersham Pharmacia Biotech, UK) using chemiluminescence according to the manufacturers recommendations. The detection reaction is based on the generation of an acridium ester by the enzymatic action of HRP on Lumigen PS-3 acridan substrates. The esters react with peroxide under slightly alkaline conditions to produce a high-intensity chemiluminescence with emission wavelength of 430 nm. The washed membranes were incubated in the chemiluminescent reagents for 5 min at room temperature in the dark. Excess reagents were drained off and the membrane wrapped in plastic. The membrane was placed in a X-ray film cassette and a sheet of Kodak Biomax autoradiography film (Kodak) was placed on top of the membrane and exposed from 15 sec to 30 min. The film was developed as described in Chapter 2, section 2.2.12.2.

4.2.6) Enzyme assays.

Wild type and mutant forms of the bifunctional PfAdoMetDC/ODC and monofunctional PfAdoMetDC and PfODC activities were determined as described in Chapter 3, section 3.2.7. Spermidine synthase activity was determined as described in Haider *et al.* (Personal communication). Results are the mean of three independent experiments performed in duplicate and expressed as a percentage of the normalised wild type controls.

4.3) RESULTS.

4.3.1) Explanations for the bifunctional nature of the PfAdoMetDC/ODC.

One explanation for the bifunctional nature of PfAdoMetDC/ODC is to allow substrate channelling to occur. For this to be true another enzyme, spermidine synthase, is required to use the decarboxylated products of PfAdoMetDC/ODC as substrate to produce spermidine. PfAdoMetDC/ODC was isolated and allowed to interact with separately expressed spermidine synthase. After co-incubation of the separately isolated enzymes for 30 min at 4°C, the proteins were analysed by SE-FPLC, followed by SDS-PAGE and activity analyses of various fractions. Fig. 4.1 indicates the size-exclusion elution profile of the interacting PfAdoMetDC/ODC and spermidine synthase. PfAdoMetDC/ODC protein and activity was observed at ~ 330 kDa, the size of the wild-type bifunctional protein. Spermidine synthase protein and activity did not co-elute with the decarboxylase activities but eluted at the expected size of ~75 kDa for the active dimeric form of spermidine synthase (Fig 4.1 B). None of the protein activities were present in fractions corresponding to a complex between PfAdoMetDC/ODC and spermidine synthase of ~404 kDa. It therefore seems that no interactions occur between PfAdoMetDC/ODC and spermidine synthase under the *in vitro* conditions used or that interactions are transient and not stable enough to survive size exclusion chromatography.

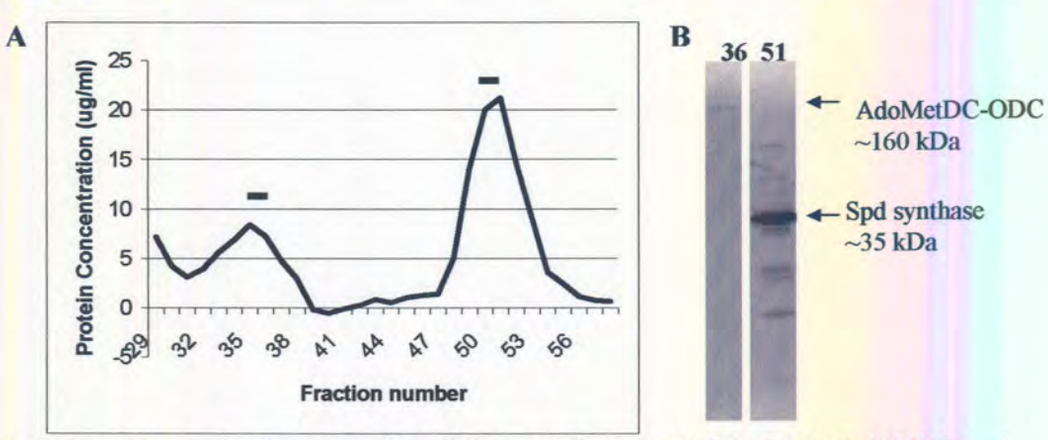

Figure 4.1: Interaction assay between the wild type bifunctional PfAdoMetDC/ODC and spermidine synthase. (A) Size exclusion elution profile of interacting PfAdoMetDC/ODC and spermidine synthase with the corresponding activities indicated in horizontal bars; PfAdoMetDC/ODC activity in fraction 36 and spermidine activity in fraction 51. **(B)** SDS-PAGE analyses of size exclusion fractions 36 and 51 corresponding to denatured protein sizes of ~160 kDa and ~35 kDa for the bifunctional, heterodimeric PfAdoMetDC/ODC and monomeric spermidine synthase, respectively.

4.3.2) Parasite-specific regions in PfAdoMetDC/ODC.

The parasite-specific inserts in the bifunctional PfAdoMetDC/ODC were defined based on a multiple-alignment of the PfAdoMetDC/ODC sequence with the corresponding sequences for the individual enzymes as described in Chapter 3 (Fig. 3.12). Briefly, the insert in the AdoMetDC domain includes residues 241-410 (insert A_1) and the ODC domain inserts include residues 1047-1085 (O_1) and 1156-1301 (O_2). The hinge region was defined by Müller *et al.* (2000) as residues 573-752 (H) (Fig. 4.2).

The identification of bifunctional AdoMetDC/ODC in other *Plasmodia* described in Chapter 3, provides the opportunity to better define the parasite-specific inserts. The large inserts in both domains, A_1 and O_2, show large variations in sequence composition and length between the three *Plasmodium* species (Fig. 4.2). The AdoMetDC domains of the murine parasite enzymes are ~100 residues longer than the *P. falciparum* enzyme. In contrast, the ODC domain is longer in *P. falciparum* due to a ~26 residue longer insert O_2. The amino acid compositions of both the large inserts (A_1 and O_2) seem to be conserved between the murine sequences but differ from *P. falciparum*, specifically in the distribution of Asn and $(NND)_x$-repeats in the *P. falciparum* sequence. The hinge region is also smaller in the murine bifunctional enzymes. One exception to the abovementioned characteristics is the smallest insert O_1 in the ODC domain, which is better conserved between the *Plasmodia* species both in terms of sequence composition and length. This suggests that this area might have a more defined function compared to the other larger, more variable inserts.

Figure 4.2: Multiple-alignment of the amino acid sequences of the bifunctional PfAdoMetDC/ODC indicating the parasite-specific areas. The putative PfAdoMetDC, hinge and PfODC domains are indicated. Amino acids shown in black boxes are >80% conserved and >60% conserved residues are shown in grey boxes. The parasite-specific inserts are in italics in the *P. falciparum* sequence (A_1: 214-410; H: 573-752; O_1: 1047-1085 and O_2: 1156-1301). Horizontal bars indicate low-complexity areas in PfAdoMetDC/ODC.

4.3.3) Sequence and structure analyses of the parasite-specific regions.

The defined parasite-specific areas of specifically the bifunctional AdoMetDC/ODC of *P. falciparum* were analysed for various sequence and structural properties to obtain information on their possible functions. The parasite-specific areas in PfAdoMetDC/ODC are rich in charged residues, predominantly Asn, Asp, Lys, Ser, Glu, Leu, and Ile (Fig. 4.2). Noticeably, both the large inserts (A_1 and O_2) and the hinge region in PfAdoMetDC/ODC are Asn-rich with insert A_1 containing 16%, the hinge region 14.8% and insert O_2, 28.7%. However, the smallest insert in the PfODC domain (O_1) is composed of 20.5% Lys residues. The inserts in PfAdoMetDC/ODC are also characterised by highly recurrent repeats of (NND)x-motifs and long stretches of Asn.

Further attempts to investigate the potential function of these parasite-specific inserts included secondary structure predictions. Kyte and Doolittle hydrophobicity analyses of the parasite-specific inserts indicated that all the inserts have a more pronounced hydrophilic nature (Fig. 4.3). Analyses of the antigenic properties of the parasite-specific regions predict several but not highly significant antigenic regions (Fig. 4.3) (Hopp-Woods equation, (Geourjon, *et al.*, 1991). The Wootton and Federhen algorithm (SEG algorithm, (Wootton and Federhen, 1996) predicted low-complexity areas in all the inserts and the hinge region except insert O_1 in the PfODC domain (Fig. 4.2 and 4.3). Furthermore, these inserts are predicted to be nonglobular with a tendency towards unstructured loops connected in the majority of cases with β-sheets (Fig. 4.3). Insert O_1 is the only area proposed to contain significant secondary structure consisting of four β-sheets arranged in an anti-parallel manner as indicated in a homology model of the PfODC domain (See Chapter 5). Thus, insert O_1 appears to be more structured without antigenic properties or low-complexity regions compared to the other parasite-specific inserts indicative of a specific role in PfAdoMetDC/ODC.

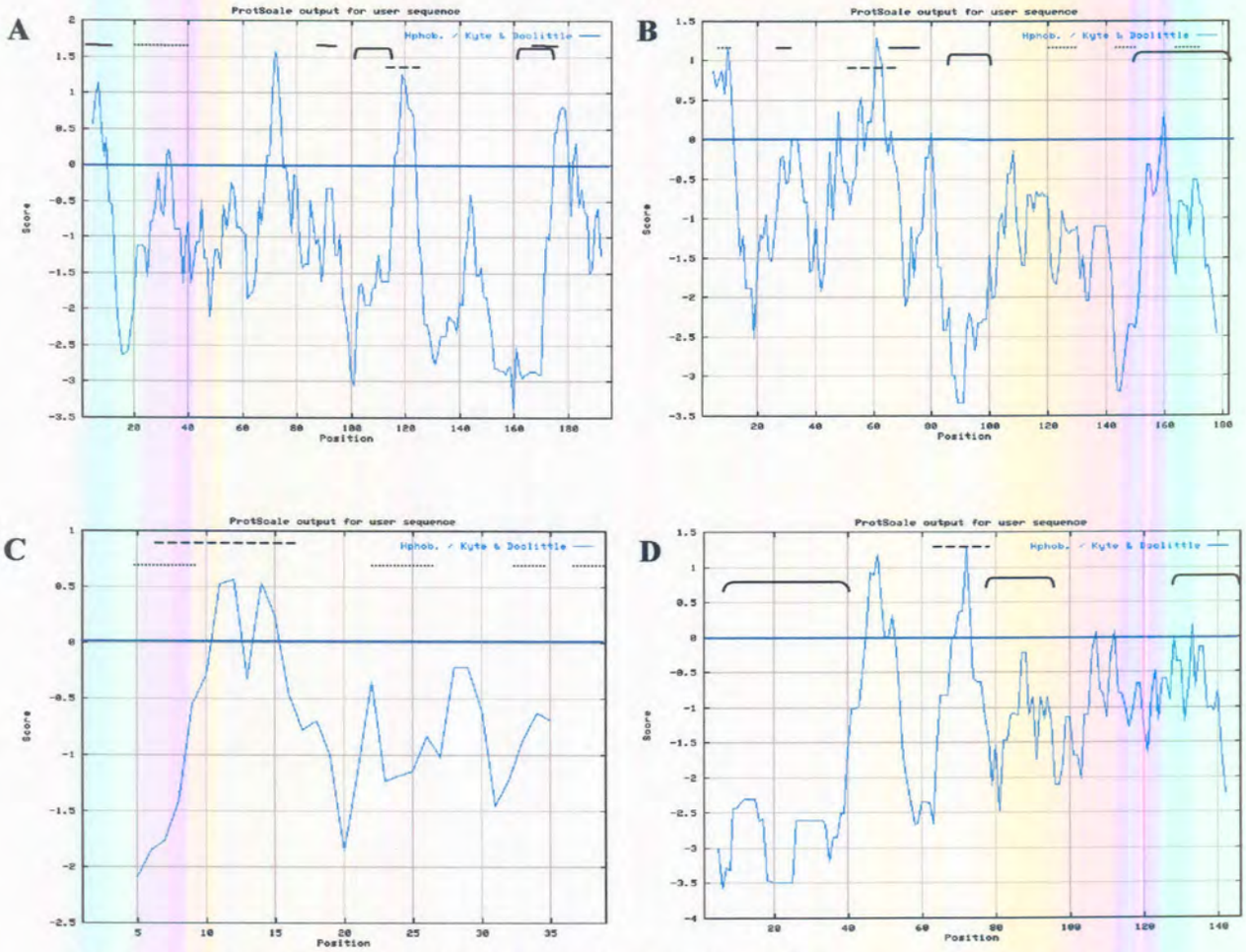

Figure 4.3: Sequence and secondary structure analyses of the parasite-specific inserts in the bifunctional PfAdoMetDC/ODC. (A) AdoMetDC insert A_1; **(B)** Hinge region; **(C)** ODC insert O_1 and **(D)** ODC insert O_2. Kyte and Doolittle hydropathy plots are shown for all four parasite-specific regions, hydrophilic residues are scored negatively. Low complexity areas are indicated with horizontal brackets. Secondary structures are predicted as α-helices (——) and β-sheets (·····) with the rest of the areas predicted to be random loops. Highest scoring antigenic regions are indicated with – – .

4.3.4) Deletion mutagenesis of parasite-specific regions in PfAdoMetDC/ODC.

To investigate what functions, if any, the parasite-specific inserts have on the activities and interactions of the bifunctional PfAdoMetDC/ODC, deletion-mutants were created for all three identified inserts as well as for the hinge region connecting the two decarboxylase activities. The mutagenic primers used to create the deletion mutants are summarised in Table 4.1.

Table 4.1: Mutagenic mega-primer oligonucleotides used for deletion mutagenesis of parasite-specific regions in PfAdoMetDC/ODC. Sites where deletion occurred are indicated with -. The sizes of the deletions (in nucleotides) are indicated.

Primer	Sequence 5'-3'	Length of primer	Tm	Number of nucleotides deleted
ΔA_1 sense	GAC GGA TAT AGC TTC TAC GTT T-AA TGA ATT TTA TTT TAC ACC TTG TGG	48	69.4	591
ΔA_1 antisense	CCA CAA GGT GTA AAA TAA AAT TCA TT-A AAC GTA GAA GCT ATA TCC GTC	48	69.4	591
ΔH sense	GTG TAG AAA AAG AAA CTT TG-G AAA AAA TGA AAG ATT ATA TAA GTG	45	54.6	540
ΔH antisense	CAC TTA TAT AAT CTT TCA TTT TTT C-CA AAG TTT CTT TTT CTA CAC	45	54.6	540
ΔO_1 sense	GGA GGG GGA TAT CCA GAA GAA TTA GAA TAT GAT-AGT TTT GAA AAA ATA TCA TTG GC	56	57.8	117
ΔO_1 antisense	GC CAA TGA TAT TTT TTC AAA ACT-ATC ATA TTC TAA TTC TTC TGG ATA TCC CCC TCC	56	57.8	117
ΔO_2 sense	GAC CAT TAC GAT CCT TTA AAT TTT T-TC TCA TAT TAT GTA AGC GAT AGT ATA TAT GG	56	69.4	435
ΔO_2 antisense	CCA TAT ATA CTA TCG CTT ACA TAA TAT GAG AA-A AAA TTT AAA GGA TCG TAA TGG TC	56	69.4	435

Each of the parasite-specific areas was individually removed from the bifunctional enzyme to create four deletion-mutants by using standard mega-primer deletion PCR techniques (Mutants A-$O_\Delta A_1$, A-$O_\Delta H$, A-$O_\Delta O_1$, A-$O_\Delta O_2$, Fig. 4.4).

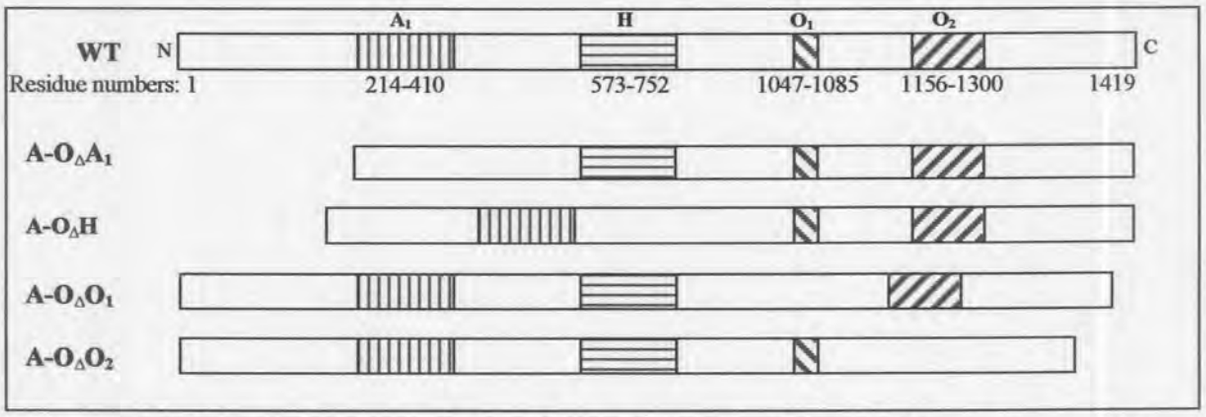

Figure 4.4: Schematic representation of the strategy used for deletion of the parasite-specific inserts and hinge region in the bifunctional PfAdoMetDC/ODC. Wild-type PfAdoMetDC/ODC is shown (top) with the positions and residue numbers of the specific inserts and the various deletion mutants indicated.

The effects of deletions on the expression of the different mutant proteins were determined by expression as for the wild-type protein followed by their isolation and analyses on SDS-PAGE (Fig. 4.5). The expressed mutant proteins had the expected decreased molecular mass. Wild-type bifunctional PfAdoMetDC/ODC migrates at the subunit size of ~160 kDa under denaturing conditions whereas mutant A-$O_\Delta A_1$ migrates at ~138 kDa corresponding to the expected size with the 21 kDa insert removed, A-$O_\Delta H$ at ~141 kDa (19.8 kDa removed), A-$O_\Delta O_1$ at ~156 kDa (4.2 kDa removed) and A-$O_\Delta O_2$ at ~144 kDa (15.9 kDa removed). The deletions did not have a major influence on the

levels of protein obtained except for the hinge deletion mutant, where expression levels dropped by 40%. This indicates that the proteins probably retained their conformations and were not expressed as misfolded proteins present in inclusion bodies in the *E. coli*.

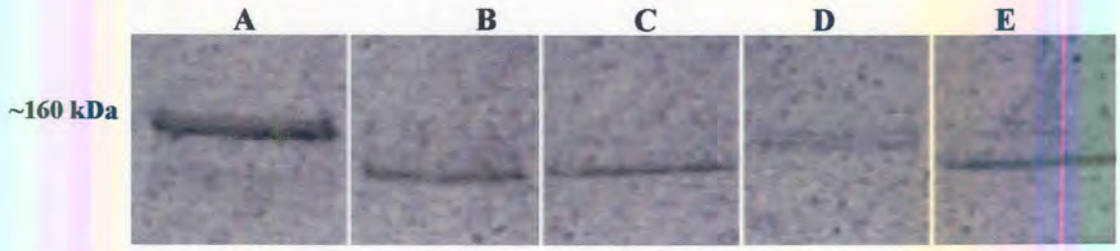

Figure 4.5: SDS-PAGE analysis of the wild-type PfAdoMetDC/ODC and the individual deletion mutants. Wild-type PfAdoMetDC/ODC (A) is compared to deletion mutants of the parasite-specific inserts (B: A-$O_\Delta A_1$; D: A-$O_\Delta O_1$ and E: A-$O_\Delta O_2$) and the hinge region (C: A-O_ΔH). Proteins were revealed with silver staining.

4.3.5) Effect of deletion mutagenesis on the decarboxylase activities of the bifunctional PfAdoMetDC/ODC.

The effects of the deleted parasite-specific inserts on the decarboxylase activities were assayed in all the mutants (Fig. 4.6). AdoMetDC and ODC activities in the bifunctional enzyme are markedly reduced (>95 %) when the deletion of the specific insert occurs inside the respective domain. Interestingly, deletion of the A_1 and O_1 inserts, both closer in linear amino acid sequence to the neighbouring domain and hinge region, also influences the activity of the neighbouring domain. For instance, deletion of O_1 reduces ODC activity by 98% (1.85 % residual activity) but also decreases AdoMetDC activity by 90% (Fig. 4.6). Deletion of A_1 reduces AdoMetDC activity by 96% and ODC activity by 72%. The effect on AdoMetDC activity is not as pronounced in the O_2 deletion mutant of ODC since AdoMetDC activity was only decreased by 47%. However, this mutant shows a 98% reduction in ODC activity. Thus, deletion of the smallest O_1 insert had the most significant effect on both enzyme activities.

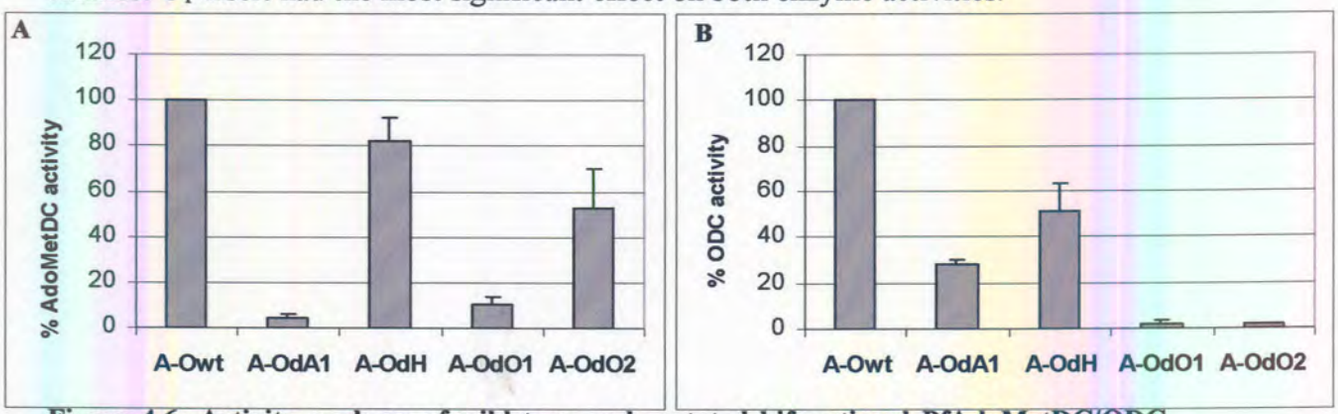

Figure 4.6: Activity analyses of wild type and mutated bifunctional PfAdoMetDC/ODC. **(A)** AdoMetDC activities and **(B)** ODC activities. Results are the mean of three independent experiments performed in duplicate and specific activities are represented as a percentage of the wild type activity with standard deviations indicated.

Deletion of the presumably flexible hinge region connecting the two domains should force the two domains into close physical proximity of each other. The PfAdoMetDC activity is only slightly decreased (18%) in this mutant but half of the PfODC activity is lost, indicating a more pronounced dependency of the ODC domain on the hinge region (Fig. 4.6). Interestingly, expression of the monofunctional PfODC domain without the hinge region also leads to a marked decrease in specific activity of the protein (Krause, *et al.*, 2000). The results presented here indicate that the parasite-specific inserts mediate specific physical interactions between the two domains that are ultimately reflected in a decrease of both decarboxylase activities.

4.3.6) Deletion mutagenesis of the parasite-specific regions in the monofunctional PfAdoMetDC and PfODC domains.

As described in Chapter 3, the individual monofunctional PfAdoMetDC and PfODC domains can be stably expressed as a heterotetrameric protein of ~145 kDa (PfAdoMetDC) and obligate homodimeric PfODC of ~170 kDa (Krause, *et al.*, 2000; Wrenger, *et al.*, 2001). The direct contribution of the parasite-specific inserts to the activities of the individually expressed monofunctional domains was investigated with single and combined deletion mutants of all the inserts in the separate domains (Fig. 4.7). The ODC domain without the hinge region was previously analysed as described in Krause *et al.* (2000). This mutant was shown to be 50% less active with a 3.4 times decreased Km for ornithine (Krause, *et al.*, 2000).

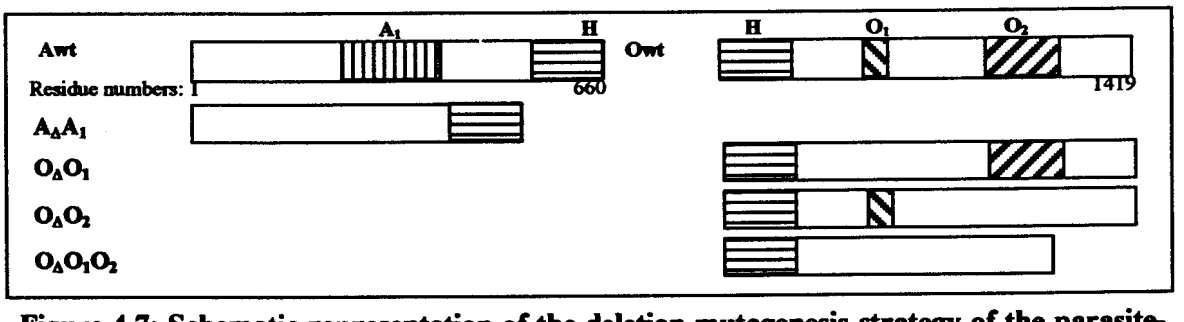

Figure 4.7: Schematic representation of the deletion mutagenesis strategy of the parasite-specific inserts in the monofunctional PfAdoMetDC and PfODC. Wild-type monofunctional PfAdoMetDC and PfODC is shown (top) with the positions of the specific inserts (A_1, H, O_1 and O_2) and the description of various deletion mutants at the bottom. Residue numbering as for the bifunctional enzyme complex.

As expected, all deletion mutants of the parasite-specific inserts did not have significant residual decarboxylase activity (Fig. 4.8). Mutant $A_\Delta A_1$ resulted in 88% decrease of AdoMetDC activity compared to 95% decrease of mutant A-$O_\Delta A_1$ in

PfAdoMetDC/ODC (Fig. 4.6). Even more pronounced activity loss was evident in the ODC domain deletion mutants: $O_{\Delta}O_1$ decreased ODC activity by 91%, whereas the ODC activities of $O_{\Delta}O_2$ and the double-deletion mutant $O_{\Delta}O_1O_2$ were reduced by 97% (Fig. 4.8). Deletion of insert A_1 in the monofunctional PfAdoMetDC domain, and double deletion of both the O_1 and O_2 inserts in the monofunctional PfODC domain results in an amino acid sequence and length that closely resemble homologues for these proteins in other organisms (See also Chapter 5). The inactivity of these mutants ($A_{\Delta}A_1$, and $O_{\Delta}O_1O_2$) implies that these areas have parasite-specific properties that are required for enzyme activity.

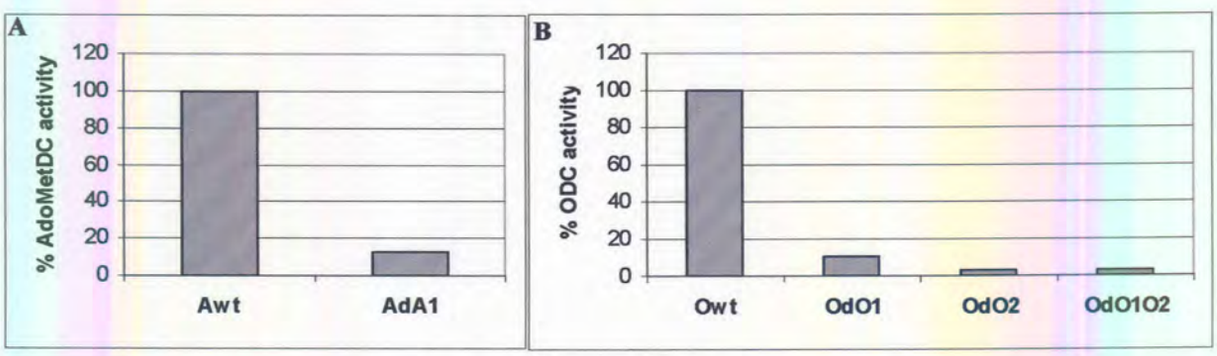

Figure 4.8: Specific activities of deletion mutants of the individual monofunctional PfAdoMetDC and PfODC domains. (A) Wild-type and deletion mutant AdoMetDC and in **(B)** Wild-type and deletion mutants activities of the ODC domain. Results are representative of duplicate experiments and specific activities are represented as a percentage of the wild type activity.

4.3.7) Oligomeric state of deletion mutants of PfAdoMetDC/ODC.

The ability of the deletion mutants to still form the heterotetrameric bifunctional complex was also investigated. The deletion mutants of the bifunctional enzyme were isolated by affinity chromatography, combined and then subjected to SE-FPLC to distinguish between heterotetrameric and heterodimeric states.

The isolated wild-type PfAdoMetDC/ODC and deletion mutants fractions were assayed for the presence of the protein (Coomassie detection and dot-blot Western) as well as for decarboxylase activity. Wild-type PfAdoMetDC/ODC elutes at a molecular mass of ~ 330 kDa (fraction 35) corresponding to a heterotetrameric complex of the AdoMetDC/ODC polypeptide as described in Chapter 3, section 3.3.5.2, Fig. 3.10 (A). Both AdoMetDC and ODC activities were observed in fractions 34-36.

Inability of the mutants to assemble in a bifunctional complex would result in the ~160 kDa heterodimeric form of the polypeptide (Fig. 4.9). Deletion of the parasite-specific

inserts in the bifunctional PfAdoMetDC/ODC did not alter the complex forming ability of the enzymes. Mutants A-$O_\Delta A_1$, A-$O_\Delta O_1$ and A-$O_\Delta O_2$ were still able to form heterotetrameric complexes with activities present only in the fractions corresponding to their predicted sizes of 276 kDa, 282 kDa and 312 kDa, respectively. The same is true for the hinge deletion mutant that also eluted around the expected smaller size of 288 kDa.

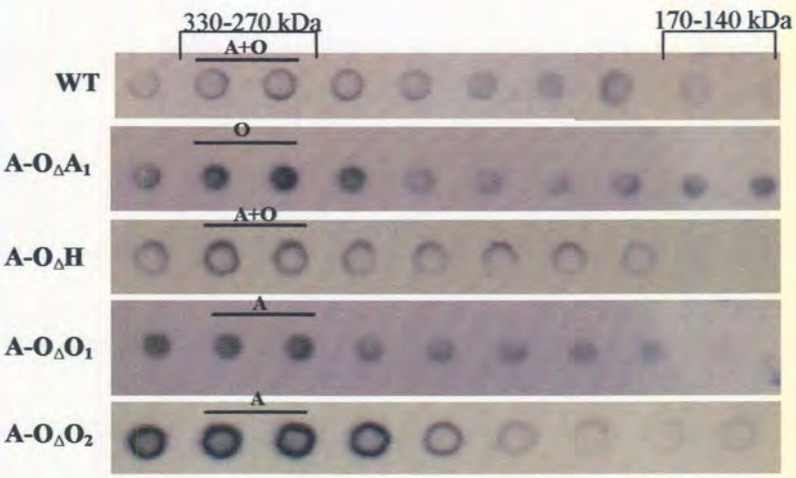

Figure 4.9: Complex forming abilities of deletion mutants of PfAdoMetDC/ODC. Protein was detected in the size-exclusion fractions with Bradford and a dot-blot Western analysis. Presence of protein and AdoMetDC and ODC activities in fractions corresponding to the heterotetrameric bifunctional complex size of 330 kDa and the uncomplexed heterodimeric form of 160 kDa is indicated. AdoMetDC and ODC activities are indicated by horizontal bars with A and O, respectively.

4.3.8) Hybrid complex forming abilities of deletion mutants of monofunctional PfAdoMetDC and PfODC.

The direct contribution of the parasite-specific inserts on hybrid bifunctional complex formation was analysed by SE-FPLC to determine the sizes of complexes after co-incubation of wild-type and mutant forms of monofunctional PfAdoMetDC and PfODC. Co-incubation of both wild-type monofunctional PfAdoMetDC and PfODC resulted in the expected formation of the ~330 kDa bifunctional heterotetrameric complex (Fig. 4.10). This is consistent with the hypothesis that the domains assembled into a heterotetramer consisting of an obligate homodimeric ODC (~170 kDa) and a heterotetrameric AdoMetDC (~145 kDa), re-establishing the natural relationship between the two domains. The new hybrid AdoMetDC + ODC complex was stable enough to survive size exclusion chromatography indicating substantial interactions between the domains.

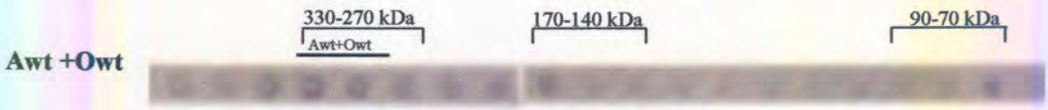

Figure 4.10: Protein-protein interactions between the separately expressed wild type AdoMetDC and ODC domains. Fractions were analysed at the different sizes corresponding to the heterotetrameric bifunctional complex size (330-270 kDa), the homodimeric size of PfODC or the heterotetrameric PfAdoMetDC (170-140 kDa) and the monomeric forms of the domains (90-70 kDa). Activities present in the different fractions are indicated in horizontal bars.

Both AdoMetDC and ODC activities were observed in the hybrid complex indicating that the association did not influence the catalytic capacity of these proteins and probably mimic the natural state of the complex (Table 4.2). Residual AdoMetDC activity was however also evident in fractions corresponding to either the heterotetrameric or heterodimeric forms of the protein.

Wild-type AdoMetDC was co-incubated with either the ODC hinge ($O_\Delta H$) or with the $O_\Delta O_1$ deletion mutants whereas wild-type ODC was co-incubated with the AdoMetDC deletion mutant $A_\Delta A_1$. Deletion of insert O_2 in the ODC domain of the bifunctional PfAdoMetDC/ODC did not seem to have a marked influence on the activity of the AdoMetDC domain and its ability to interact with AdoMetDC was not further investigated. Fig. 4.11 indicates the individual activities and dot-blot analyses of the various size exclusion fragments of the abovementioned combinations.

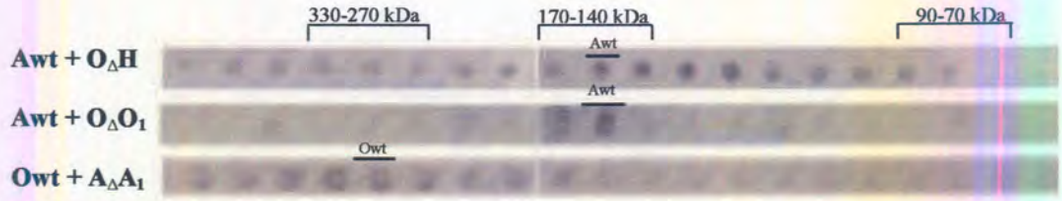

Figure 4.11: Intermolecular interaction between the wild-type and mutant forms of the monofunctional AdoMetDC and ODC. Fractions were analysed at the different sizes corresponding to the heterotetrameric bifunctional complex size (330-270 kDa), the homodimeric size of PfODC or the heterotetrameric PfAdoMetDC (170-140 kDa) and the monomeric forms of the domains (90-70 kDa). Dot blot analyses of the interaction of the wild type decarboxylase domains as well as interactions between AdoMetDCwt (Awt) and $O_\Delta H$; Awt and $O_\Delta O_1$ and ODCwt (Owt) and $A_\Delta A_1$ after size exclusion chromatography. Activities present in the different fractions are indicated in horizontal bars.

Physical association still occurred between the wild-type ODC domain and the AdoMetDC mutant $A_\Delta A_1$ to form a heterotetrameric bifunctional complex of ~330 kDa, indicating that the contribution of insert A_1 to bifunctional complex formation is

probably not that pronounced. The majority of the AdoMetDC activity after removal of the hinge region from the ODC domain (mutant $O_{\Delta}H$) was associated with the heterotetrameric form (~145 kDa, Table 4.2). Very little AdoMetDC activity and no ODC activity was detectable in a bifunctional complex and low levels of ODC activity was detected in fractions corresponding to the homodimeric form of the protein (~180 kDa, Table 4.2). Deletion of the much smaller O_1 insert in the ODC domain also resulted in active heterotetrameric PfAdoMetDC (~145 kDa) with no AdoMetDC activity in a bifunctional complex. Table 4.2 summarises the hybrid complex formation abilities of the mutant proteins.

Table 4.2: Hybrid complex formation abilities of mutant forms of the monofunctional AdoMetDC and ODC. Hybrid bifunctional complexes are found in the 330-270 kDa size range, monofunctional heterotetrameric AdoMetDC and homodimeric ODC in the 180-140 kDa range and monomeric proteins in the 90-70 kDa range. Results are indicated as a combination of the presence of protein (detected with the dot-blot assay) with the number of crosses indicating the relative enzyme activities. nd: Not detectable.

	330-270 kDa		180-140 kDa		90-70 kDa	
	AdoMetDC	ODC	AdoMetDC	ODC	AdoMetDC	ODC
Awt + Owt	+++	+++	++	nd	++	nd
Awt + $O_{\Delta}H$	+	nd	+++	+	++	nd
Awt + $O_{\Delta}O_1$	nd	nd	+++	nd	+	nd
Owt + $A_{\Delta}A_1$	nd	+++	nd	++	nd	nd

Hybrid bifunctional complex formation is only possible with the co-incubation of wild-type ODC and the AdoMetDC deletion mutant $A_{\Delta}A1$. No physical interactions seem to occur between the domains when the hinge region or insert O_1 is removed from the ODC domain. These areas therefore seem to enable formation of the hybrid bifunctional complex.

4.4) DISCUSSION.

4.4.1) Explanations for the bifunctional nature of the PfAdoMetDC/ODC.

Various proposals were forwarded to explain the bifunctional nature of several proteins of *P. falciparum* including substrate channelling, intramolecular communication and coordinated regulation of transcription and translation (Müller, *et al.*, 2001). In the case of DHFR-TS, substrate channelling has been proposed since the two proteins catalyse consecutive reactions in the same metabolic pathway (Miles, *et al.*, 1999). For this to be true for the bifunctional PfAdoMetDC/ODC it should interact with the subsequent

enzyme in the polyamine biosynthetic pathway, spermidine synthase, to allow the production of spermidine. Size exclusion analyses indicated that no physical association occurs between the proteins under the *in vitro* conditions used. It therefore appears unlikely that substrate channelling can be forwarded as an explanation for the bifunctional organisation of the PfAdoMetDC/ODC. More advanced techniques like the yeast two-hybrid system need to be employed to further investigate interactions with other proteins. Hypotheses advanced thus far for the bifunctional nature of PfAdoMetDC/ODC include facilitation of coordinated regulation of the protein levels and activities of both proteins or intramolecular communication and interaction.

4.4.2) Defining the parasite-specific inserts in PfAdoMetDC/ODC.

A characteristic feature of many *P. falciparum* proteins is the tendency towards large gene coding regions containing different sized inserts and increased protein sizes relative to homologues from other organisms (Bowman, *et al.*, 1999; Gardner, *et al.*, 1998). Parasite-specific inserts is usually defined with multiple amino acid sequence alignment which shows that these areas separate mutually-conserved blocks when compared to other homologous proteins (Chapter 3, Fig. 3.12 for PfAdoMetDC/ODC). Three different parasite-specific inserts were identified in PfAdoMetDC/ODC, a single large insert in the AdoMetDC domain and two inserts in the ODC domain as described in Chapter 3. The large inserts in both domains show extensive sequence composition and length variability between the different *Plasmodium* species, the AdoMetDC insert shows 18% sequence identity and the large insert in the ODC domain, O_2, shows 27% identity. However, the smallest insert in the ODC domain is more conserved (O_1, 46% identical). This suggests that the variability of this insert is constrained due to some, as yet undetermined, function.

Multiple alignment of the PfAdoMetDC/ODC amino acid sequence with homologues from other organisms as well as the bifunctional forms of the proteins in the other *Plasmodia* highlighted the inherent difficulties in predicting the correct boundaries of the hinge region between the two domains. The *L. donovani* ODC sequence has an ~200 residue extension at the N-terminus but is not part of a bifunctional protein (Hanson, *et al.*, 1992). The hinge region of the bifunctional AdoMetDC/ODC in contrast is much shorter in the sequences of the two other *Plasmodia*. Furthermore, removal of the hinge region leads to a 50% reduction in the activity of monofunctional PfODC (Krause, *et al.*, 2000). It is therefore possible that the actual hinge region in *P. falciparum* is smaller

than the currently defined 180 residues. Part of the current hinge region could therefore constitute partial sequence of ODC. This possibility is supported by the results of the deletion mutagenesis studies discussed in detail in the following sections.

4.4.3) Structural properties of the parasite-specific regions in PfAdoMetDC/ODC.

The precise function and evolutionary advantage of parasite-specific inserts in *P. falciparum* proteins are not known. Proposed functions include interaction sites with as yet undefined regulatory proteins in the parasite and sites for interaction with host proteins as a means to evade the host immune response (Li and Baker, 1998; Schofield, 1991). Analyses of the sequence and predicted structural properties of the identified parasite-specific inserts in PfAdoMetDC/ODC indicated that the inserts and the hinge region are rich in charged residues. Asp- and Asn-rich areas have been characterized in other *Plasmodial* proteins, particularly in antigenic regions of membrane proteins and as such may play a role in the diversity of the parasite population to evade the hosts defence mechanisms by acting as antigenic smokescreens (Barale, *et al.*, 1997; Kemp, *et al.*, 1987; Reeder and Brown, 1996). However, the inserts in PfAdoMetDC/ODC do not show significant antigenicity (Fig. 4.3) and it is furthermore a cytosolic protein (Müller, *et al.*, 2000). As mentioned in Chapter 3, such Asp-Asn rich areas are also found in repeat areas in proteins that would normally not be exposed to the host immune system (Coppel and Black, 1998; Schofield, 1991). Asn/Glu-rich areas have also been described as 'prion-domains' that function during protein-protein interactions and link functional domains in certain proteins (Michelitsch and Weissman, 2000; Wickner, *et al.*, 2000). Perutz has suggested that poly-Glu or poly-Asn repeats and possibly regions rich in other polar residues might behave as modular mediators of protein-protein interactions termed 'polar zippers' because of the capacity of their side chains to form hydrogen bonded networks (Perutz, *et al.*, 1994). Removal of the parasite-specific insert in glutathione reductase led to an unstable protein indicating that these areas are important in the folding and stability of the protein (Gilberger, *et al.*, 2000). Therefore, in the absence of distinctive, predicted antigenic properties of the inserts, it is possible that the parasite-specific areas containing Nx and (NND)x-repeats in PfAdoMetDC/ODC might be involved in protein-protein interactions to stabilise the heterotetrameric bifunctional complex.

In Chapter 3, the complete amino acid sequence of the bifunctional PfAdoMetDC/ODC was analysed for possible low-complexity regions, the majority of which were located

in the parasite-specific inserts and hinge region (Fig. 3.12 and 4.2). This correlates with the results of Pizzi *et al.* who showed that such areas found in hydrophilic regions in *P. falciparum* proteins co-inside with parasite-specific, rapidly diverging insertions (Pizzi and Frontali, 2001). The low-complexity regions found in more conserved areas of the protein make up a minor subset of prevalently hydrophobic areas and are proposed to be involved in the core structure of these proteins (Pizzi and Frontali, 2001).

4.4.4) Involvement of the parasite-specific inserts in the decarboxylase activities of PfAdoMetDC/ODC.

It is thus possible that the parasite-specific inserts function as interaction sites to enable formation of the bifunctional PfAdoMetDC/ODC protein or catalytic activities. These possible functions were investigated by determination of the effects of deletion of the parasite-specific inserts on bifunctional complex formation and domain activities.

Deletion mutagenesis of the parasite-specific inserts in the bifunctional PfAdoMetDC/ODC indicated that the inserts are essential for the activity (and inherent conformation) of the involved domain (Fig. 4.6). However, deletion of the parasite-specific inserts also affect the activity/conformation of the neighbouring domain. Specifically, the inserts closer to the neighbouring domain in terms of the linear amino acid sequence (A_1 and O_1) seem to have the greatest influences on the activity of the other domain. Previous point mutation studies of the active site residues indicated that the two decarboxylase activities in PfAdoMetDC/ODC are able to function independently (Wrenger, *et al.*, 2001). The results presented here indicate that the parasite-specific inserts are involved in specific communication between the two domains in the bifunctional complex.

All the deletion mutants in the covalently linked PfAdoMetDC/ODC were still able to form bifunctional complexes. It is possible that the parasite-specific inserts do not act alone in the stabilisation of the bifunctional protein complex but that other interactions also contribute to the stabilisation and conformation of the individual domain. Cumulative interactions have been proposed to stabilise the *T. brucei* ODC dimeric interface (Myers, *et al.*, 2001). It was furthermore shown that mutation of single residues far removed from the active site decreased catalytic activity mostly due to long-range energetic coupling of these residues to the active site (Myers, *et al.*, 2001). Therefore, deleting the parasite-specific inserts in PfAdoMetDC/ODC may display

similar effects on catalytic activities of the individual domains even though their properties indicate a surface location far removed from the actual active site centres.

4.4.5) Characterisation of the physical association between the decarboxylase domains.

The physical association between AdoMetDC and ODC was confirmed by the stable reassociation of the individually expressed monofunctional PfAdoMetDC and PfODC domains in a bifunctional hybrid complex that reflects the properties of the bifunctional PfAdoMetDC/ODC (Fig. 4.10). This species-specific physical contact seems to be mediated in part by a parasite-specific insert in the ODC domain. Insert O_1 is predicted as a structured hydrophilic area that does not contain low-complexity regions and shows minimal sequence variability between *Plasmodium* species (Fig. 4.2 and 4.3). This region is also more important for both decarboxylase activities in the bifunctional PfAdoMetDC/ODC (Fig. 4.6). The larger parasite-specific inserts do not seem to mediate physical interactions between the two domains (Fig. 4.11 and Table 4.2). These areas show large sequence length and property diversion between different *Plasmodial* species (Fig. 4.2 and 4.3). Structural analyses of the large parasite-specific inserts (A_1 and O_2) predicted hydrophilic, nonglobular regions containing low-complexity areas in agreement with the results of Pizzi and Frontali (Pizzi and Frontali, 2001). These authors showed that similar inserts in other *Plasmodial* proteins are also hydrophilic in nature and contain low-complexity regions. These low-complexity regions were proposed not to be involved in the correct folding of the proteins and most probably form nonglobular domains that are extruded from the protein core. This might illustrate that if large global conformation changes occurred due to the deletion of such large areas (such as A_1 and O_2) it did not influence enzyme activity as much as deletion of a smaller, more structured insert (O_1). Further investigations are required to determine the exact role of the low-complexity areas within these parasite-specific inserts in complex formation and activities of the two domains. The effect of deletions of inserts in the ODC domain on the formation of the obligate homodimer is not known at this stage. It is not unlikely that these deletions may prevent formation of the homodimeric state thereby preventing association with AdoMetDC.

Deletion of the hinge region in PfAdoMetDC/ODC had a more pronounced impact on ODC activity in the bifunctional enzyme (50% reduced) than on AdoMetDC activity (18% reduced). Monofunctional ODC lacking the hinge region is only half as active as

ODC expressed with part of the hinge region (Krause, *et al.*, 2000) and prevents the ODC domain to interact with the wild type AdoMetDC domain. It therefore seems that the hinge region is important for the stability/activity of the ODC domain by mediating the correct folding of the domain to ensure the active homodimeric ODC or by actually constituting part of the ODC domain and is smaller than the currently defined 180 residues.

None of the deletion mutants of the bifunctional PfAdoMetDC/ODC exhibit any effects on heterotetrameric complex formation (Fig. 4.9). However, in contrast to the results obtained with the deletion mutants of the bifunctional complex, deletions of the hinge region or insert O_1 in monofunctional ODC prevented formation of the hybrid, heterotetrameric AdoMetDC/ODC complex. However, deletion of insert A_1 in monofunctional AdoMetDC had no effect on hybrid complex formation. Taken together the results obtained with the deletion mutants of the bifunctional and monofunctional enzymes suggest that the AdoMetDC insert is not involved in heterotetrameric complex formation but only in protein-protein interactions affecting its own activity and that of the ODC domain. The roles of the ODC inserts seem to be more complex since their deletion affects not only its own activity and that of AdoMetDC but also heterotetrameric complex formation. Furthermore, when the complex formation of the mutated bifunctional proteins is compared to the hybrid complex formation of the mutant monofunctional proteins, it seems possible that the complex formation in the bifunctional proteins were due mostly to interactions between the AdoMetDC domains, with no apparent contribution of the mutant ODC. However, the ODC domain is more refractory to change to be able to interact with the AdoMetDC domain when these monofunctional proteins are not covalently linked.

The data presented here indicate that although the two decarboxylase activities can function independently of each other, physical protein-protein interactions are present in the bifunctional PfAdoMetDC/ODC that has effects on both enzyme activities and heterotetrameric complex formation. Future investigations on the role of the parasite-specific inserts in these protein-protein interactions include: 1) Deletion of only the low-complexity areas in the parasite-specific inserts to determine their contribution to protein-protein interactions and validating their proposed surface locality; 2) Mutation of only the polar residues (e.g. Asp and Asn repeats) in the parasite-specific inserts to apolar residues. This should provide evidence for the polar zipper theory of modular

mediators of protein-protein interactions and furthermore limit global conformational changes due to deletion of the large, parasite-specific inserts; 3) Determination of the oligomeric state of the mutated monofunctional proteins to determine the weight of the contribution of each domain to the protein-protein interactions via their ability or not to still form their own intramolecular interactions.

This chapter contributed to understanding the structure-functional relationships that stabilise the bifunctional heterotetrameric PfAdoMetDC/ODC. In Chapter 5, detailed structural characteristics of a three-dimensional homology model of the ODC component of the bifunctional enzyme are described.

CHAPTER 5

Comparative properties of a homology model of the ornithine decarboxylase component of the *P. falciparum* S-adenosylmethionine decarboxylase/ornithine decarboxylase.

5.1) INTRODUCTION

The understanding, modification and manipulation of protein function generally require knowledge of the three-dimensional (3D) structure of a protein at the atomic level. Detailed knowledge of the structure and function of the individual ODC and AdoMetDC enzymes of the malaria parasite is required in order to clarify and understand their arrangement and interactions in the unique bifunctional complex. Structural data is available for the murine, human, *T. brucei* and *Lactobacillus* ODC enzymes (Almrud, *et al.*, 2000; Grishin, *et al.*, 1999; Kern, *et al.*, 1999; Momany, *et al.*, 1995; Vitali, *et al.*, 1999) and for the human AdoMetDC enzyme (Ekstrom, *et al.*, 1999) but not for the malarial enzymes.

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the preferred methods to obtain detailed 3D structural information of proteins but both methods are time consuming and require large quantities of protein. In addition, NMR spectroscopy cannot resolve the structures of proteins larger than 200 residues or those of flexible proteins, while X-ray crystallography depends on the generation of suitable crystals (Sali, 1995). The high A+T content of the parasite genome contributes in many instances to the low or insignificant expression of protein from heterologous systems (Baca and Hol, 2000; Chang, 1994). Furthermore, crystallization of malarial proteins is problematic due to the characteristic prevalence of regions of low-complexity and/or inserted amino acids, as evidenced by the paucity of protein crystal structures.

Comparisons between various proteins have demonstrated that their tertiary structures are usually better conserved in evolution than their amino acid sequences (Blundell, *et al.*, 1987; Srinivasan, *et al.*, 1996). It is generally accepted that high sequence similarity is reflected by distinct structure similarity. The root mean square deviation (RMSD) for

protein α-carbon backbones sharing 50% primary sequence identity is expected to be better than 1 Å. This served as the premise for the development of knowledge-based comparative protein structure modelling methods by which a homology model for a new protein is extrapolated from the known three-dimensional structure of related proteins (Fig. 5.1) (Peitsch, 1996; Sali, 1995). This has resulted in the application of comparative or homology modelling as an additional and/or alternative method to obtain protein structural information (Srinivasan, *et al.*, 1996).

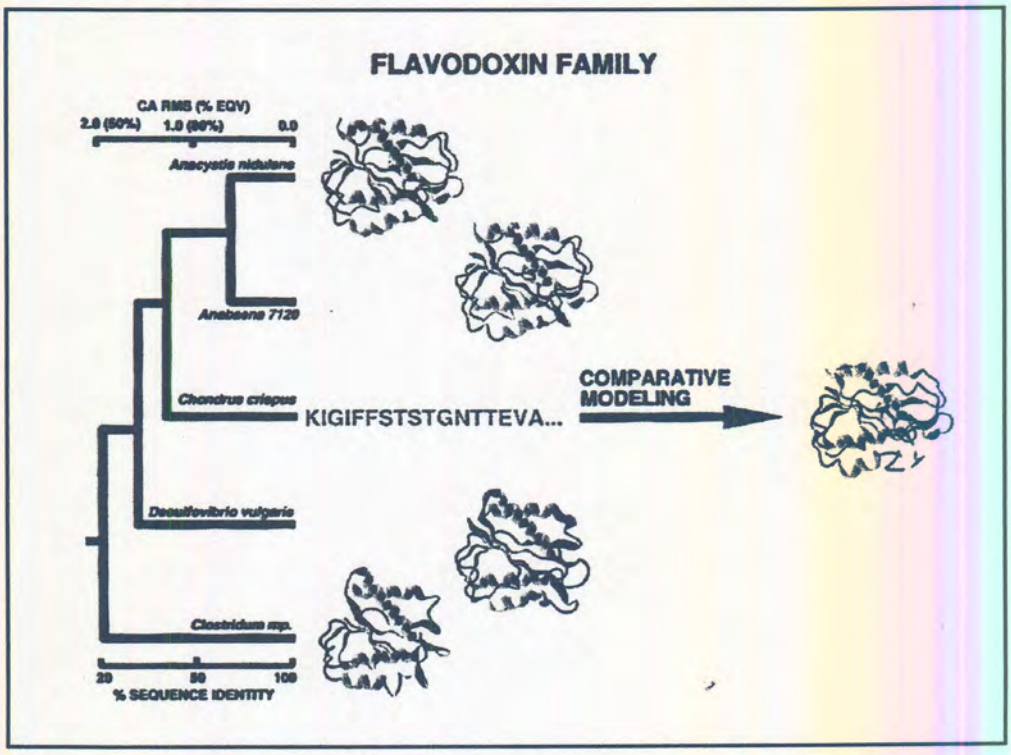

Figure 5.1: Comparative homology modelling due to the evolutionary precept that protein families have both similar sequences and 3D structures. The flavodoxin family is depicted where a protein in this family can be modelled from its sequence using the other structures in the family. The tree shows the % sequence and structural similarity. Adapted from (Sali, 1995).

A recent version of the Protein Information Resource Protein Sequence Database (PIR-PSD 71.03) contained 283 138 entries of protein sequences on 15th of February 2002. In contrast, the Protein Data Bank of experimentally determined protein structures contained only 17 304 structures on the 12th of February 2002. Since about one third of known sequences appear to be related to at least one known structure, the number of sequences that can be modelled is an order of magnitude larger than the number of experimentally determined protein structures (Oregano, *et al.*, 1994). Furthermore, the usefulness of homology modelling is increasing because the various genome projects are producing more sequences and the speed at which novel protein folds are being

determined is increasing due to the application of high-throughput methods (Sanchez, *et al.*, 2000; Taylor, 2002).

Homology modelling uses experimentally determined protein structures as templates to predict or extrapolate the conformation of another protein that has a similar amino acid sequence (>40% identity) (Sanchez and Sali, 1997). This is possible because a small change in the sequence usually results in a small change in the 3D structure. Insertions and deletions occurring during evolution are usually confined mainly to loops between secondary structures and do not alter the fold of the protein. This gives rise to families of protein folds having related structures but varying sequence identities (Jones, *et al.*, 1996). Homology modelling consists of four sequential steps beginning with the identification of the proteins with known 3D structures that are related to the target sequence and used as templates for the structure extrapolation (Srinivasan, *et al.*, 1996)(Fig. 5.2). The second and most crucial step is to align these sequences with the target sequence. This is followed by building of the model for which various calculations are possible. In the fourth step, the model is evaluated using a variety of criteria. These steps can be repeated until a satisfactory model is obtained as indicated in Fig. 5.2. The main difference between the various comparative modelling methods is in how the 3D model is calculated from the sequence alignment (Srinivasan, *et al.*, 1996). The original method of modelling used rigid-body assembly to model a protein from a few core regions, loops and sidechains obtained from related structures. The rigid bodies are assembled on a framework defined as the average of the α-carbon atoms in the conserved regions of certain folds. Another method uses modelling by segment matching that relies on the approximate positions of conserved atoms from the templates to calculate the coordinates of other atoms. The third group of methods involves modelling by satisfaction of spatial restraints with either distance geometry or optimisation techniques (Blundell, *et al.*, 1987; Sali, 1995; Srinivasan, *et al.*, 1996).

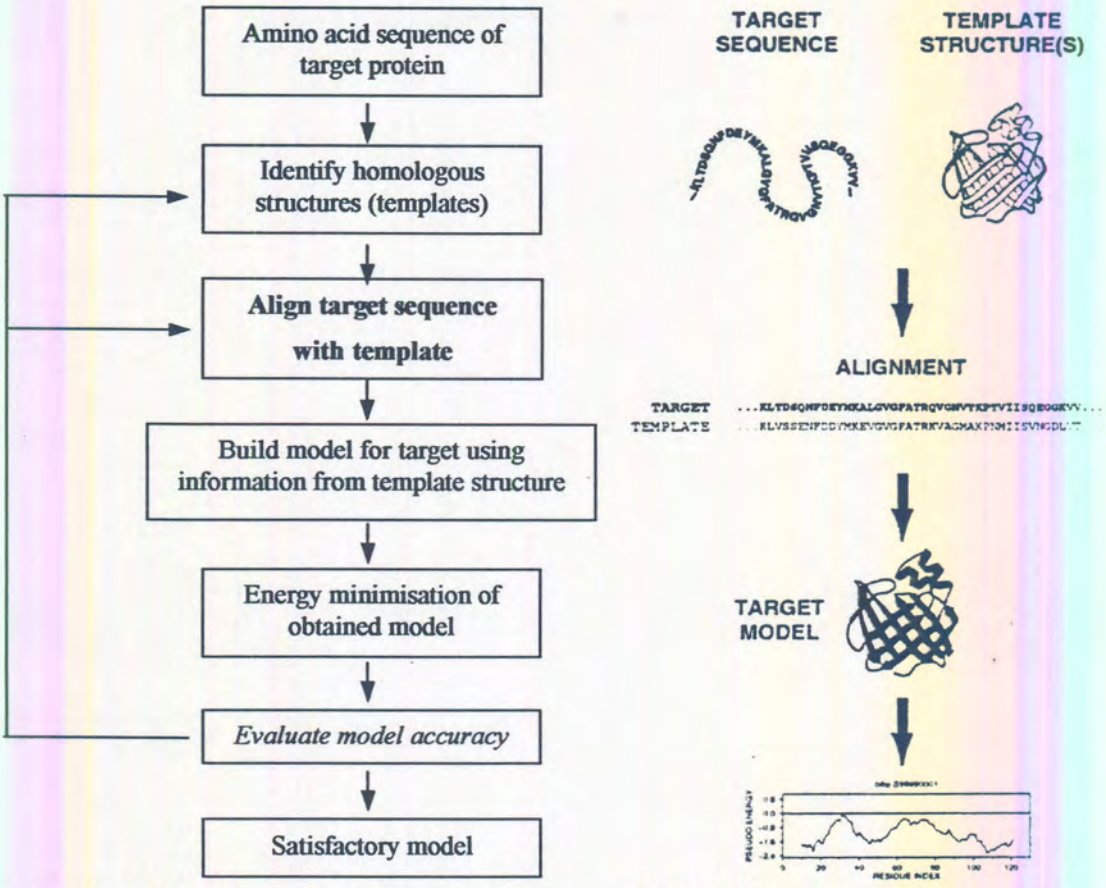

Figure 5.2: Steps in comparative protein structure modelling. Adapted from (Srinivasan, *et al.*, 1996).

No crystal structure could be obtained yet for malarial dihydrofolate reductase (DHFR) notwithstanding its widespread resistance to anti-folates and importance as a selective and validated antimalarial drug target. Malarial DHFR also occurs in a bifunctional complex with thymidylate synthase (TS) and contains inserted amino acids (Lemcke, *et al.*, 1999; Toyoda, *et al.*, 1997). The mechanism of resistance of DHFR to known antimalarials however, could be explained by a homology model and furthermore led to the discovery of lead inhibitors (Lemcke, *et al.*, 1999; Rastelli, *et al.*, 2000; Toyoda, *et al.*, 1997 Warhurst, 1998 #151).

The AdoMetDC structure for the human enzyme has been crystallised to 2.25 Å resolution (Fig. 5.3 A (Ekstrom, *et al.*, 1999). The mature protein is a dimer consisting of two α- and β-chains. The architecture of each αβ monomer is a novel four-layer α/β-sandwich fold, comprised of 2 antiparallel 8-stranded β-sheets flanked by several α- and 3_{10} helices (Ekstrom, *et al.*, 1999). The low sequence identity (<20%) between the

AdoMetDC domain of the bifunctional PfAdoMetDC/ODC (excluding the parasite-specific regions) and AdoMetDCs from other organisms (Chapter 3) complicates comparative structural modelling of this protein.

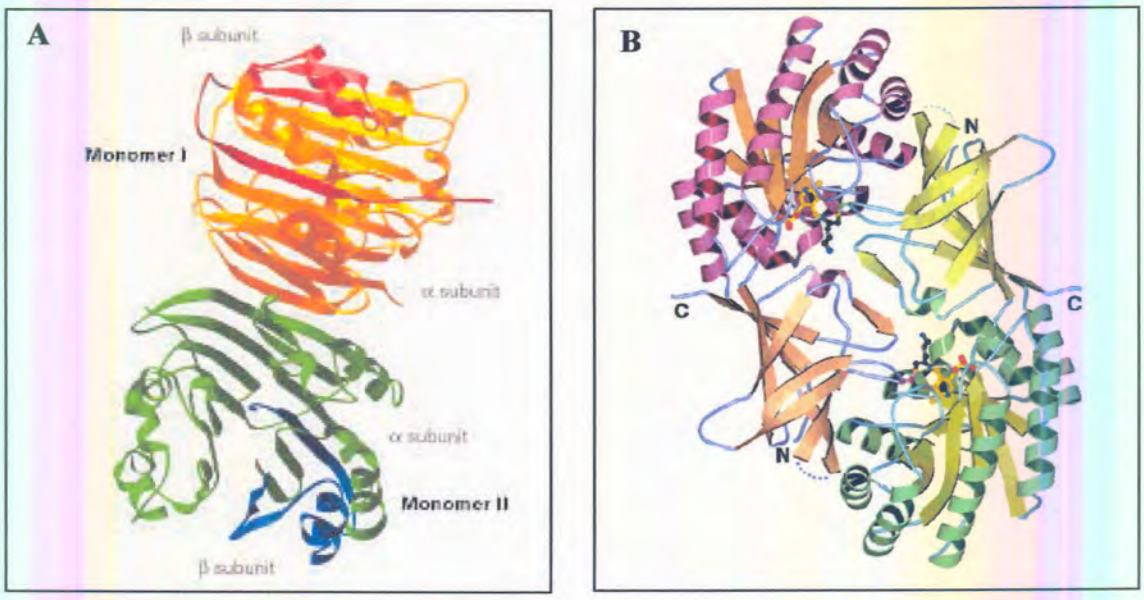

Figure 5.3: Crystal structures of mammalian AdoMetDC and protozoal ODC. (A) The AdoMetDC dimer from *H. sapiens* (Ekstrom, *et al.*, 1999) and in **(B)** the ODC structure from *T. brucei* (Grishin, *et al.*, 1999).

ODC from *T. brucei*, *H. sapiens* and *M. musculus* have been crystallised (Almrud, *et al.*, 2000; Grishin, *et al.*, 1999; Momany, *et al.*, 1995). In all the eukaryotic ODC structures, the protein is found as a dimer with each monomer consisting of two distinct domains, a N-terminal α/β-barrel and a C-terminal β-barrel domain (Fig. 5.3 B). The contacts at the dimer interface are primarily in the C-terminal domains of each monomer, whereas the active site is found in the barrel of the N-terminal domain and is closed-up by a loop from the C-terminal domain of the second monomer projecting into the cavity. The extent of sequence identity (~30%) between ODCs from a variety of species and the malarial ODC domain of the bifunctional PfAdoMetDC/ODC (excluding the parasite-specific regions, Chapter 3) suggested the feasibility of a comparative protein structure modelling approach for the generation of its three-dimensional structure.

This chapter describes the generation of a homology model of only the PfODC component of the bifunctional PfAdoMetDC/ODC and compares the putative structure to other experimentally determined ODC structures.

Some of the results obtained in this chapter have been accepted for publication in Proteins: Structure, Function and Genetics (Birkholtz, *et al.*, 2002a). The results have also been presented as papers at international meetings (Birkholtz, 2000e; 2002c; Birkholtz, *et al.*, 2000a) and as posters at international and national conferences (Birkholtz, 2000b; 2000c; Birkholtz, *et al.*, 2001b; 2001c).

5.2) MATERIALS AND METHODS

5.2.1) *In silico* analyses of predicted structural motifs of PfODC.

PfODC was structurally classified by different *in silico* techniques. A hierarchical classification of ODCs was obtained with the CATH database (Orengo, *et al.*, 1997). Protein domain structures are grouped according to a novel hierarchical classification, which clusters proteins at four major levels, Class (C), Architecture (A), Topology (T) and Homologous superfamily (H). The class is derived from secondary structure content and is automatically assigned for more than 90% of protein structures. Architecture, which describes the gross orientation of secondary structures, independent of connectivities, is currently assigned manually. The topology level clusters structures according to their topological connections and numbers of secondary structures. The homologous superfamilies cluster proteins with highly similar structures and functions. The assignments of structures to topology families and homologous superfamilies are made by sequence and structure comparisons. SCOP (Murzin, *et al.*, 1995) allows the Structural Classification of Proteins by providing a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structures are known. As such, it provides a broad survey of all known protein folds and detailed information about the close relatives of any particular protein. Comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences. The DALI server (http://www.ebi.ac.uk/dali/) (distance matrix alignment) allows the comparison of the 3D protein structure to all the other known structures of proteins in the Protein Data Bank (Holm and Sander, 1993). Low complexity regions in PfODC were identified with the SEG Program (Wootton and Federhen, 1996).

5.2.2) Comparative modelling of monomeric PfODC.

Comparative homology modelling was performed according to the method originally described by Browne in 1969 (Blundell, *et al.*, 1987; Browne, *et al.*, 1969). This is based on the assembly of a number of rigid bodies obtained from aligned protein structures followed by framework calculations, generation of mainchain atoms of the core regions, generation of loops and modelling of the sidechains based on their intrinsic conformational preferences and the conformation in the template structures. The stereochemistry of the model is then improved with energy minimisations.

Multiple pairwise alignment with the CLUSTAL W programme (Thompson, *et al.*, 1994) was used to compare the PfODC amino acid sequence as described in Chapter 4. The malaria-specific insert O_2 (residues 1139-1296) and the hinge region (residues 573-837) identified in this alignment were removed from the PfODC sequence due to the absence of the corresponding sequence in the modelling template and absence of known structural homologous. The remaining 411 residues (838-1138/1297-1406) of the ODC component was submitted to the SWISS-Model server (Automated Protein Modelling Server, Version 3.5, GlaxoWellcome Experimental Research, Geneva, Switzerland; (Guex, *et al.*, 1999; Guex and Peitsch, 1997; 1999) for comparative protein structure modelling by rigid body assembly with the following knowledge-based approach (Peitsch, 1995a; Peitsch, 1995b; Peitsch, 1996; Peitsch, *et al.*, 1996): Suitable templates on which to base the model were found by searching all similarities within the target sequence compared to sequences of known structures, using BLASTP2 searches of the ExNRL-3D database (SWISS-Model sequence database, reflecting the protein sequences of ExPDB). Templates with a sequence identity above 25% and larger than 20 residues were selected by SIM© and used to detect domains that could be modelled based on unrelated templates (Huang and Miller, 1991). ProModII was subsequently employed to generate models using ExPDB (The structure database used by the SWISS-Model is derived from the Brookhaven Protein Data Bank (PDB, BNL). Energy minimization and structure refinement was done with GROMOS96 to reduce steric overlap specifically in side-chains (default parameters using steepest gradient for 200 cycles with Gromos96 force field, BIOMOS b.v. Company).

The resulting model was validated manually with the WHAT_CHECK module of the WHAT IF program (version 19970813-1517; (Vriend, 1990) and with the PROCheck program (Laskowski, *et al.*, 1993). Molecular surfaces and potentials were created with

GRASP (Graphical representation and analyses of structural properties; Columbia University, New York; (Nicholls, *et al.*, 1991).

Models were visualized and edited with SWISS-PDB Viewer and analysed with the InsightII package (Accelrys, San Diego, USA) on a Silicon Graphics Octane workstation (Silicon Graphics, Mountainview, USA). The SWISS-PDB Viewer scenes were rendered with POV-Ray.

5.2.3) Dimerisation of monomeric PfODC.

The dimeric form of PfODC was built by superimposing the PfODC monomers on the dimeric *T. brucei* crystal structure using the Improved fit module of SWISS-PDB Viewer and merging the coordinates into one planar field. The resulting dimeric PfODC was then subjected to energy minimization with the Discover3 module of the InsightII package (cff91 force field for 10 000 iterations with a conjugate gradient) and checked for any disallowed bumps occurring between the two different chains. Interacting residues were analysed with Protein Explorer (http://www.umass.edu/microbio/chime/explorer) and LigPlot (Version4.0; (Wallace, *et al.*, 1995). The structure was analysed for accuracy with the WHAT_CHECK module of the WHAT IF program (Vriend, 1990).

5.2.4) Docking of ligands into the active site of dimeric PfODC.

Active site residues were identified as those corresponding to proven functional residues in the active site pockets of the *T. brucei* and human ODC crystal structures (Almrud, *et al.*, 2000; Coleman, *et al.*, 1993; Grishin, *et al.*, 1999; Osterman, *et al.*, 1994). These include Lys_{69}, Arg_{154}, His_{197}, $Gly_{235\text{-}237}$, Glu_{274}, Arg_{277}, Tyr_{389}, Asp_{332}, Cys_{360} and Asp_{361} (numbers according to the *T. brucei* protein). Structures for PLP and ornithine were generated with the Builder module of the InsightII package and their energies minimized as described above. Binding of PLP and ornithine requires the formation of a Schiff-base between the two ligands with ornithine then also forming a covalent bond to the S_γ atom of Cys_{360} (*T. brucei* numbering). In order to dock this transition state complex of PLP-ornithine into PfODC, a structure for the linked PLP-ornithine was created and allowed to form a covalent link with Cys_{1355}. The ligand-ODC complex was then minimized as described above. Possible interactions between the ligands and residues in PfODC were analysed with LigPlot (Wallace, *et al.*, 1995). The structures were analysed for accuracy with the WHAT IF program (Vriend, 1990).

5.2.5) Limited proteolysis studies.

Limited proteolysis is a powerful tool for probing the higher order structure of proteins by using classical biochemical methods (Hubbard, 1998). This is achieved under nondenaturing conditions by limiting the proteolytic reaction of various proteases through altering the reaction conditions such that digestion of every susceptible peptide bond is prevented and only the location of certain bonds with respect to the overall fold of the protein is obtained. Recombinantly expressed PfODC (ODC domain containing the hinge region, Chapter 3, section 3.2.2) was subjected to limited proteolysis according to a modification of the methods by Hubbard and Wilkinson (Hubbard, 1998; Wilkinson, 2000). Briefly, the expressed protein was isolated as described in Chapter 3, section 3.2.2.2 and subjected to either proteinase K (Roche, Mannheim, Germany) or trypsin (Macherey-Nagel, Duren, Germany) digestion. Enzyme:substrate ratios were optimised at between 1:50 and 1:100 and ~ 750 ng protein was subjected to proteolysis at room temperature in 10 mM Tris-HCl (pH 8.0) for 0, 5, 20 and 60 min intervals. The reactions were stopped by addition of 0.1 mM PMSF and 2 x SDS-PAGE loading dye and boiling for 5 min. The digested samples were then analysed on a 12.5 % SDS-PAGE and stained with silver as described in Chapter 3, section 3.2.6. Limited proteolysis sites were predicted by analysing the obtained dimeric PfODC model with the Nickpred Server (sjh.bi.umist.ac.uk/cgi-bin/npred/nickpred) (Hubbard, 1998).

5.3) RESULTS.

5.3.1) Structural classification of PfODC.

Comparisons between homologous proteins have shown that conformations are better conserved in evolution than the corresponding amino acid sequences (Srinivasan, *et al.*, 1996). The CATH database places ODC in a hierarchical fashion with the Lyase homologous superfamily that shares topologies, consisting of a barrel-like architecture, with lyases and thrombin. These proteins are grouped into the mainly β single domain class of proteins according to evolutionary and structural groupings. SCOP places ODC in a superfamily of PLP-binding proteins that include the alanine racemase-like family, based on a triosephosphate isomerase (TIM) barrel-like fold. The α/β-barrel structure found in these proteins seems extremely well preserved even in distant homologues (Alanine racemase and TIM proteins) with diverse functions.

5.3.2) Modelling monomeric PfODC.

In order to apply homology modelling, the first non-trivial step is to obtain a multiple-alignment of the query amino acid sequence against sequences from other known structures. At present, there are no known homologues of the inserted or hinge region sequences. The largest insertion of 158 residues (O_2) and the hinge region were thus removed in order to arrive at a satisfactory homology model based on multiple sequence alignments used to describe and define the inserts in Chapter 4. Subsequent pair-wise alignment of the remaining 411 residues of PfODC showed the highest identity to the amino acid sequences of the *T. brucei* enzymes (41.54 %) and 39.23 % with the mouse enzyme (PDB # 7ODC, Fig. 5.4). The crystal structure of the *T. brucei* enzyme obtained with bound co-factor PLP (PDB # 1QU4 at 2.9 Å) was therefore used as template to build the PfODC homology model. Fig. 5.4 shows the alignment between the 411 PfODC residues and *T. brucei* ODC used to create the model and also indicates the secondary structural elements for each protein as predicted by the Swiss-Model server.

The resulting homology model consisted of 373 residues based on the template structure plus the *ab initio* constructed 39 residue malaria-specific insert O_1. The root mean square of deviation (RMSD) value between the model carbon α-backbone and the *T. brucei* structure was 0.816 Å as determined by PROFIT. This value increased to 6.917 Å when the *ab initio* modelled insert O_1 of 39 residues was included in the comparison.

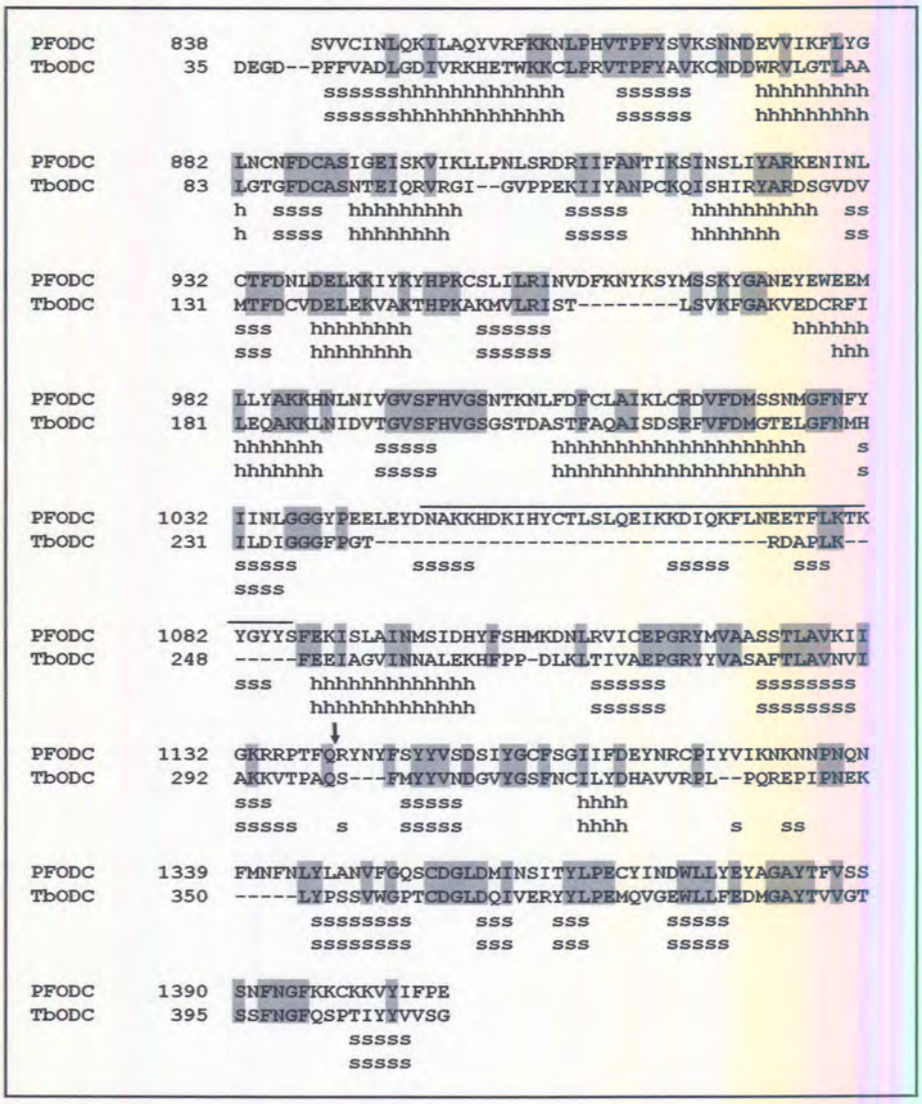

Figure 5.4: Sequence alignment of *P. falciparum* ODC (PfODC) and the template used for homology modelling, *T. brucei* ODC (TbODC, PDB: 1QU4) obtained with SIM© using default parameters. Identical residues are shaded and the secondary structural elements are indicated: s for β-sheets and h for α-helices. The site where insert O_2 was removed to create the PfODC model is indicated with an arrow. Insert O_1 is indicated with the black bar.

5.3.3) Evaluation of the PfODC model quality and accuracy.

PROCheck analyses of the PfODC model that included insert O_1 produced a Ramachandran plot in which 84.2% of the residues were in the most favoured regions indicating the model to be sterically acceptable (Fig. 5.5). The only residues that were present in disallowed areas were Leu_{1061}, Glu_{1063} and Phe_{1077}. These residues form part of insert O_1 and predicting their exact conformational arrangement is therefore impeded by the inherent difficulty in modelling unknown loops.

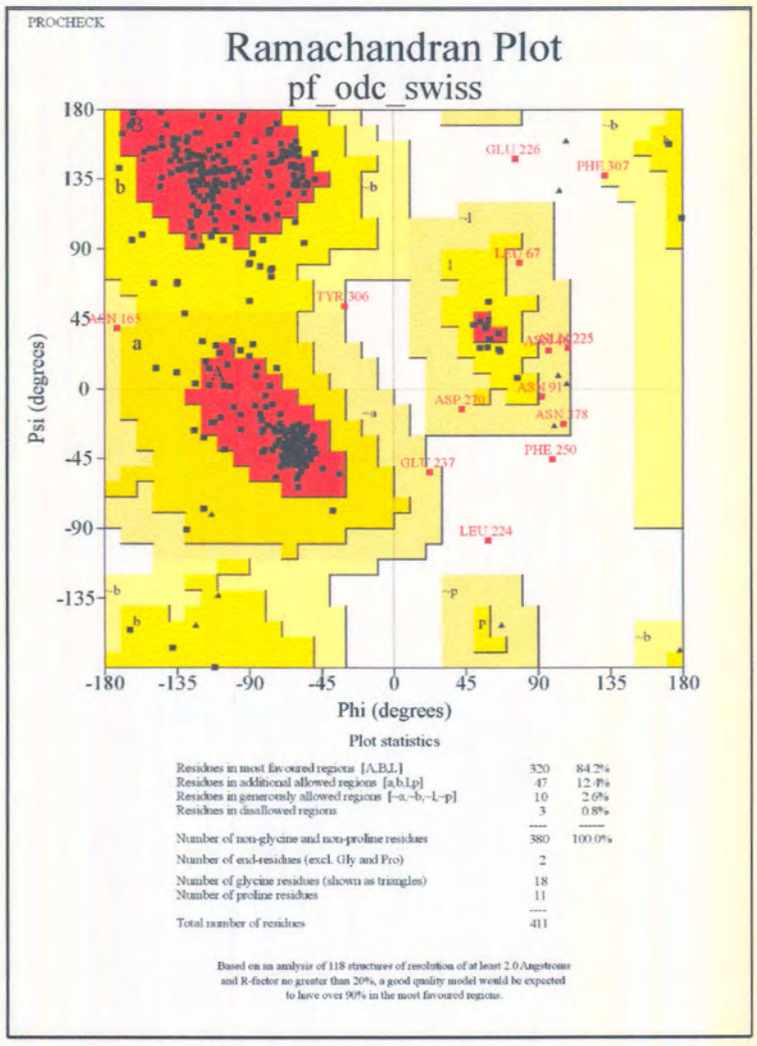

Figure 5.5: Ramachandran plot for the model of PfODC produced by PROCHECK. 84.2% of the residues are present in favourable structural areas with the exception of 3 residues in disallowed regions (Glu, Leu and Phe).

Most of the main-chain and side-chain parameters were better than typical values allowed and the rest of the parameters were within the allowed ranges (Fig. 5.6).

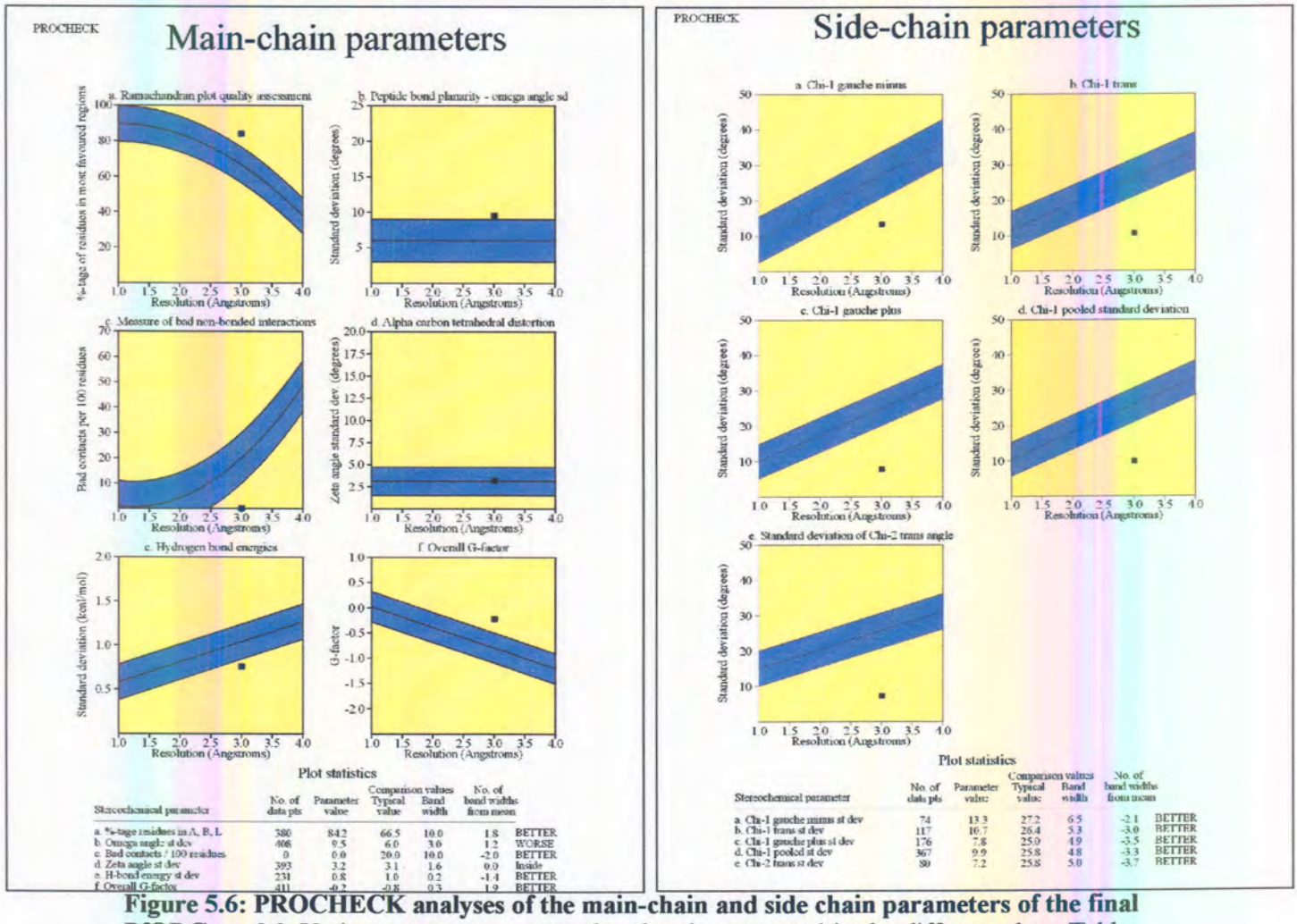

Figure 5.6: PROCHECK analyses of the main-chain and side chain parameters of the final PfODC model. Various parameters are analysed and represented in the different plots. Tables evaluate parameters into allowed ranges.

Quality analyses were also performed with the program WHAT_CHECK from the WHAT IF package and several stereochemical parameters are summarized in Table 5.1. From this Table, the quality of both the monomeric and dimeric forms of PfODC seem as good as those of the reported *T. brucei* and *H. sapiens* structures with which it is compared.

Table 5.1: Summary of WHAT IF quality assessment data. Data for the *T. brucei* (PDB # 1QU4) and *H. sapiens* (PDB # 1D7K) ODC enzymes are compared with the monomeric and dimeric form of PfODC. Values are structure Z-scores with + better than normal. * RMS Z-scores should be close to 1.0.

Structure	2nd generation packing quality	Ramachandran plot appearance	χ1/χ2 rotamer normality	Backbone conformation	Bond lengths*	Bond angle variability*	Omega angle restraints*
1QU4	-2.231	-2.051	-1.848	-0.052	0.612	1.674	1.527
1D7K	-1.1	-1.2	-0.3	-0.6	0.753	1.791	2.360
PfODC:							
Monomer	-3.658	-0.762	0.263	-1.681	1.282	1.290	1.295
Dimer	-3.978	-0.807	0.283	-1.849	1.282	1.291	1.294

5.3.4) Characterisation of monomeric PfODC.

The PfODC monomer consists of two distinct domains, a N-terminal α/β TIM barrel and a C-terminal modified Greek-key β-barrel (Fig. 5.7 A). These features correlate closely to all the other eukaryotic ODC structures characterized to date (Almrud, *et al.*, 2000; Grishin, *et al.*, 1999; Kern, *et al.*, 1999; Momany, *et al.*, 1995; Vitali, *et al.*, 1999). Evaluation of the relationships with the human ODC crystal structure (PDB # 1D7K) (Almrud, *et al.*, 2000) indicates large similarities, especially in the structural motifs of the N-terminal α/β barrel and the C-terminal β-sheet (Almrud, *et al.*, 2000) (Fig. 5.7 B). Superimposition of the PfODC monomer on the human ODC structure yields a RMSD of 0.80 Å (involving 1348 atoms) with the areas scoring the worst B-factors being the loops connecting the well-conserved structural elements of the core protein. These regions are elongated in the malaria protein compared to other ODCs. The insert O_1 is predicted to have four anti-parallel β-sheets (I1-I4) with the first two longer than the second pair and does not contain low complexity areas (Fig. 5.7 A). No firm conclusions are possible with regard to the exact orientation of the bulk of the loop in space. In the model, the insert seems to lie parallel to the rest of the protein and to bulge out towards the C-terminal domain on the same side of the monomer as the entry to the active site. No significant interactions are apparent between the protein core and the insert. The attachment site for insert O_2, which was removed in PfODC to create the model, is in the C-terminal three-quarter of the loop between C2b and C3 (Fig. 5.7 A).

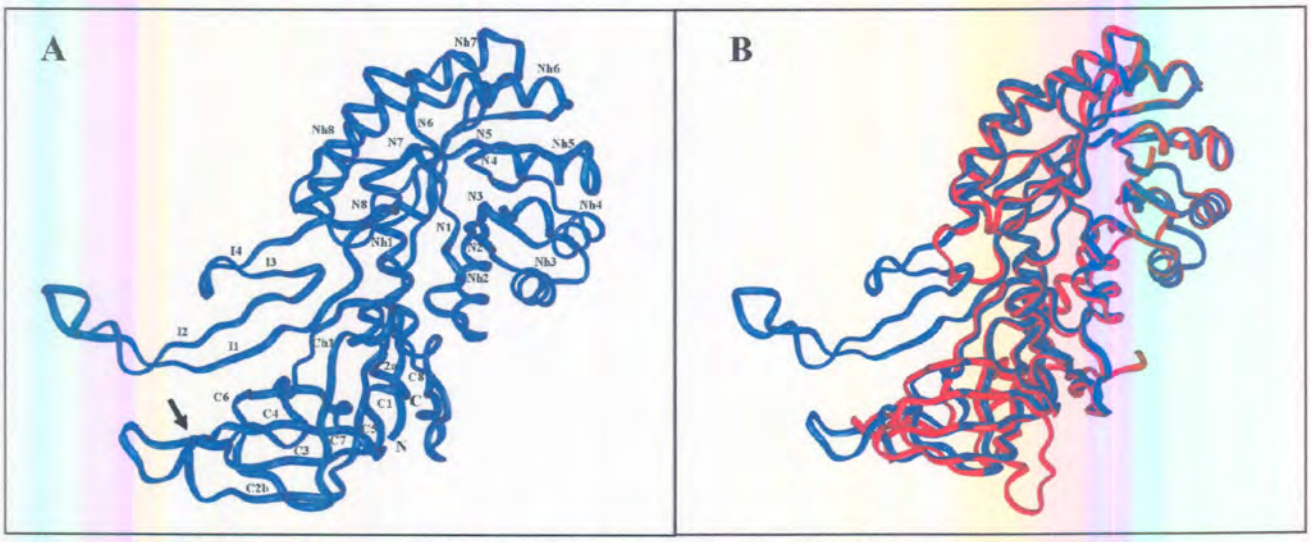

Figure 5.7: Ribbon diagram of the homology model for the PfODC monomer (A) and in (B) compared with the human enzyme. (A) The β-sheets are labelled in succession starting from the N-terminus (N1-N8, I1-I4, C1-C8) and the α-helices are specified in the N-terminal α/β-barrel domain (Nh1-Nh8). The site where insert O_2 was removed to create the PfODC model is indicated with an arrow. **(B)** PfODC in blue is superimposed on the human ODC structure in red.

5.3.5) Characterisation of dimeric PfODC.

Eukaryotic ODC is an obligate homodimeric enzyme with two active sites at the interface between the two monomers (Cohen, 1998; Pegg, *et al.*, 1994). The PfODC monomers were superimposed on the dimeric *T. brucei* crystal structure using the Improved fit module of SWISS-PdbViewer followed by energy minimization with Discover3 to create a dimeric structure of the malarial enzyme. Minimization was performed using a conjugate gradient to a maximum derivative of 0.0030 after 10 000 iterations involving 1398 atoms with RMSD of 0.37 Å and energy of –15 350 kcal/mol. Convergence to a lower derivative was not obtained, probably due to the presence of the malaria-specific areas present as unconstrained loops on the surface of the protein. The RMSD values of the structures prior to and after minimization were 2.370 Å and 3.045 Å for the backbone and side-chain atoms, respectively. Analyses of the quality of the dimer as indicated with the WHAT IF program indicates a good working structure as compared with the *T. brucei* and human structures (Table 5.1).

The dimer comprises of a head-to-tail association of the two monomers, with the C-terminal domain of one monomer vertical to the N-terminal domain of the second monomer (Fig. 5.8).

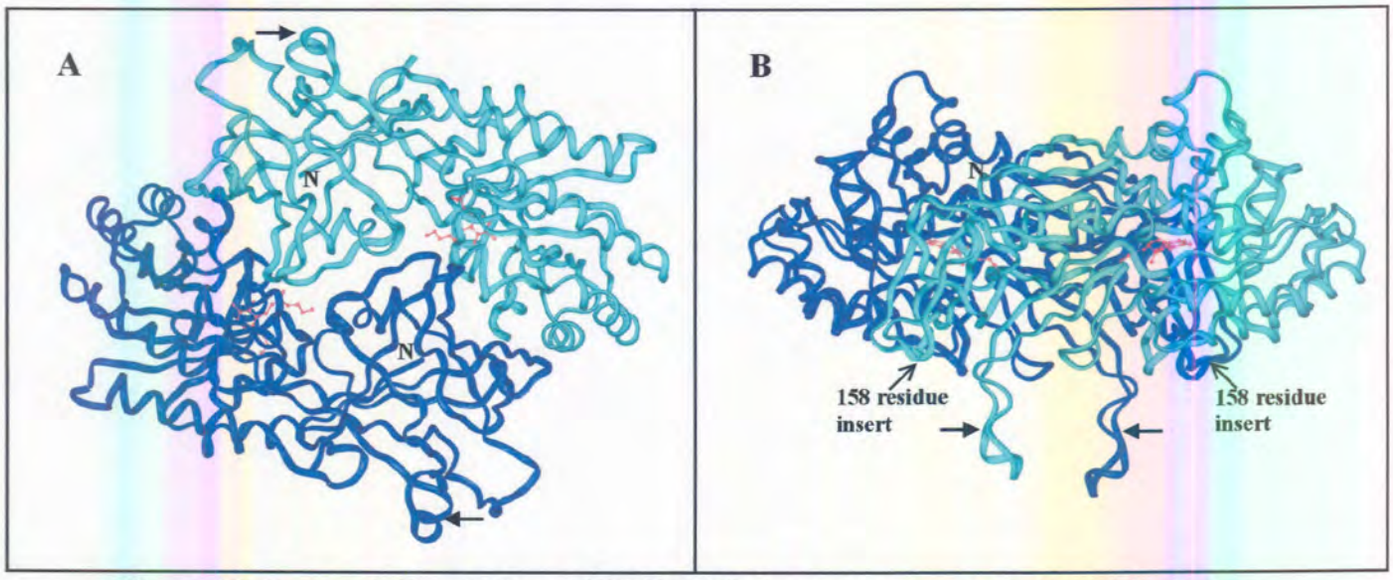

Figure 5.8: Proposed dimeric form of PfODC. The two monomers are indicated in shades of blue and the dimer is viewed from the bottom (A) and side (B). The PLP cofactor and DFMO inhibitor is indicated in red in the two active site pockets formed at the interface between the two monomers. The N-terminus in each monomer is indicated. The location of the 158 residue insert O_2 removed to create the model in the protein is shown in **(B)**. The filled arrows indicate the 39 residue insert O_1 that was modelled.

Several interactions between these two domains are apparent (Fig. 5.9). As is the case in the *T. brucei* enzyme, the dimer interface of PfODC is characterized by an aromatic amino acid zipper (Grishin, *et al.*, 1999). Phe_{1392}, Tyr_{1305} and Phe_{1319} (substituting a second Tyr residue found in *T. brucei*) are involved in hydrophobic contacts across the dimer interface forming an anti-parallel stacked interaction via their aromatic rings. A pronounced hydrophobic contact is predicted between Tyr_{1305} and Ile_{915} in an area that is well conserved in terms of sequence identity in all ODCs but not in PfODC (residues 111-115: ANPCK; PfODC residues 912-916: ANTIK). A salt bridge (1.94 Å) is predicted between Asp_{1359} and Arg_{1134} from the opposite monomer as well as between Lys_{1133} and Glu_{1369} (1.85 Å). There are several stabilizing interactions close to the active site residues. Particularly for PfODC, Lys_{970} is predicted to interact with various residues surrounding the active site Cys_{1355} donated by the opposite monomer. Lys_{970} forms a hydrogen bond with Asp_{1356} and hydrophobic contacts with Gly_{1357}, Asp_{1359} and Gly_{1352}.

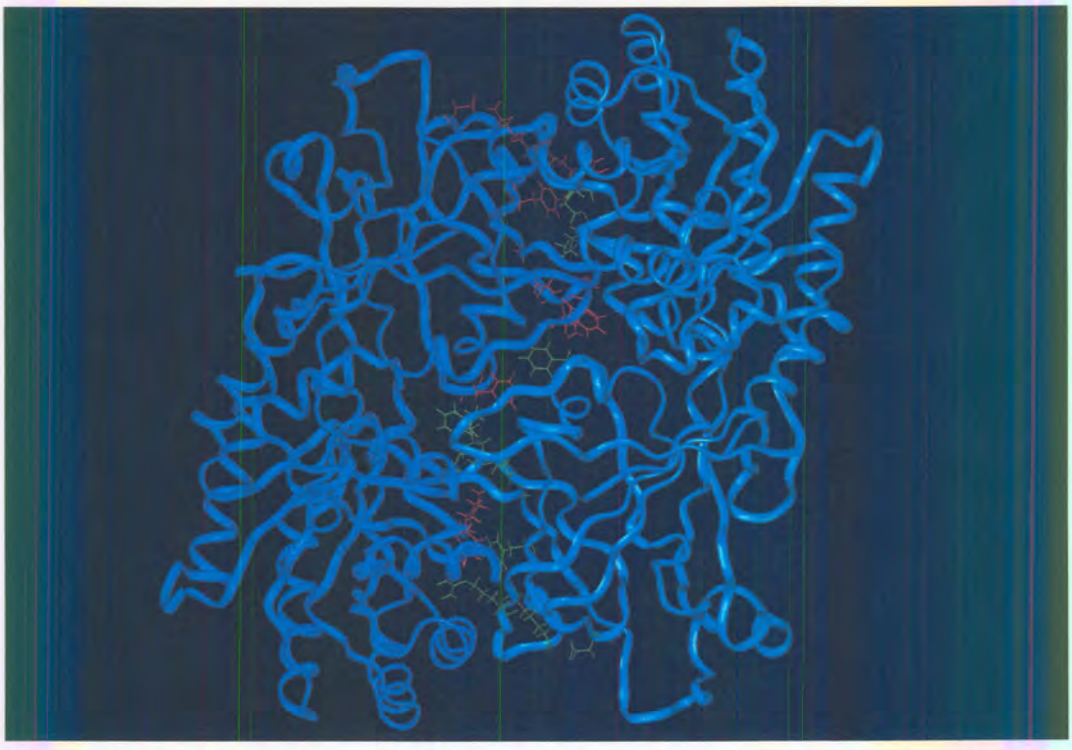

Figure 5.9: Interactions at the ODC dimer interface. The monomers are indicated in different shades of blue and the residues donated from each monomer in red and green respectively.

5.3.6) Active site pocket of dimeric PfODC.

In order to analyse the active site pocket of PfODC, the binding of PLP and ornithine was simulated. Transition state structures for PLP bound to ornithine via a Schiff-base was used and minimized into the active site pocket to an energy of –15 314 kcal/mol for

the protein-ligand complex. The predicted PLP and substrate binding site of the PfODC has a few residues within 3.0 Å to enable hydrogen bonding and charge interactions (Fig. 5.10).

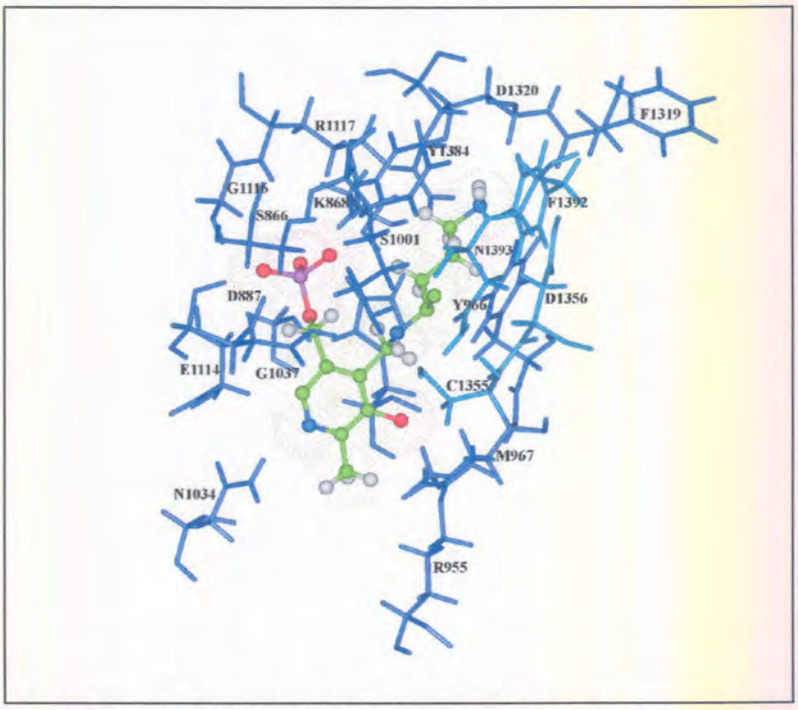

Figure 5.10: Active site residues of the PfODC indicating the interactions with PLP and ornithine. PLP and Schiff-base linked ornithine are indicated in ball-and-sticks coloured for their different atoms and with van der Waals surfaces shown. Residues are coloured in different shades of blue indicating the contribution by the two respective monomers.

Possible residues interacting with either PLP or ornithine were identified using LigPlot and are summarized in Table 5.2. The catalytic residues showed similar spatial orientations as those in the human structure. From these data it is clear that the active site residues are conserved compared to the *T. brucei* and human enzymes in terms of binding to PLP with Cys_{1355} (from the second monomer), Asp_{887}, Arg_{955}, His_{998}, Ser_{1001}, Gly_{1037}, Glu_{1114}, Gly_{1116} and Tyr_{1384}. However, residues Thr_{933}, Met_{967} and Asn_{1034} were present only in the PfODC-PLP binding site. The substrate binding site was derived from interactions with ornithine and consists of mutually conserved (compared to *T. brucei* and the human enzymes) residues Lys_{868}, Asp_{1320}, Cys_{1355}, Asp_{1356}, Tyr_{1384}, Phe_{1392} and Asn_{1393}. In the PfODC model, Cys_{1355} makes contact with Lys_{868} as well as Ala_{889} from the opposite monomer. As with the PLP-binding site, two residues (Tyr_{966} and Arg_{1117}) were only found in the PfODC substrate-binding site and not in either the *H. sapiens* or *T. brucei* binding sites. Of the five PfODC-specific residues characterising the entire active site pocket of this enzyme, only Thr_{933} and Arg_{1117} are conserved in comparison to the primary amino acid sequences of *H. sapiens*, *T. brucei* and *L.*

donovani. The identification of PfODC-specific residues Met_{967}, Asn_{1034} and Tyr_{966} could find application in the rational design of PfODC-specific lead inhibitors.

Table 5.2: Active site residues involved in interactions with ornithine as substrate and PLP as co-factor. Active site residues were identified with LigPlot v 4.0 for the *T. brucei*, *H. sapiens* and *P. falciparum* ODC enzymes. For PfODC the numbering is according to the bifunctional enzyme complex and A and B indicate which monomer contributed a residue towards the active site.

Ligand	*T. brucei* residues	*H. sapiens* residues	PfODC residues
PLP		Cys360	Cys1355A
	Arg154	Arg154	Arg955B
	Glu274	Glu274	Glu1114B
	Asp88	Asp88	Asp887B
	His197	His197	His998B
	Ala67	Ala67	Ser866B
	Tyr389	Tyr389	Tyr1384B
	Ser200		Ser1001B
	Gly276	Gly276	Gly1116B
	Arg277	Arg277	
	Gly236	Gly236	Gly1037B
	Gly237	Gly237	
			Thr933B
			Asn1034B
			Met967B
Substrate	Tyr389	Tyr389	Tyr1384B
	Cys360		Cys1355B
	Phe397	Phe397	Phe1392A
	Asp361		Asp1356A
	Tyr331		Phe1319B
	Asp332		Asp1320B
		Asn398	Asn1393A
		Val68	
		Ala67	
		Lys69	Lys868B
		Asn71	
		Glu94	
		Cys70	
		Ala392	
			Arg1117B
			Tyr966B

5.3.7) Analysis of the molecular surface of PfODC.

To obtain an indication of the surface properties of the PfODC model, the monomeric form of the model was analysed with GRASP to create the molecular surface and indicate the specific potentials of certain projections and cavities. Fig. 5.11 indicates the surfaces of PfODC (Fig. 5.11 A) compared with the surface of the *H. sapiens* ODC (Fig. 5.11 B). PfODC has overall a more positively charged surface especially in the C-

terminal β-sheet domain. Pronounced electro-negative areas are found in the α/β-barrel. One distinct difference is the electrostatic potential at the surface of the active site pockets of the PfODC model and human crystal structure. PfODC has a very pronounced positively charged ring at the entrance to the negatively charged inner pocket whereas this division is not as distinct for the human enzyme.

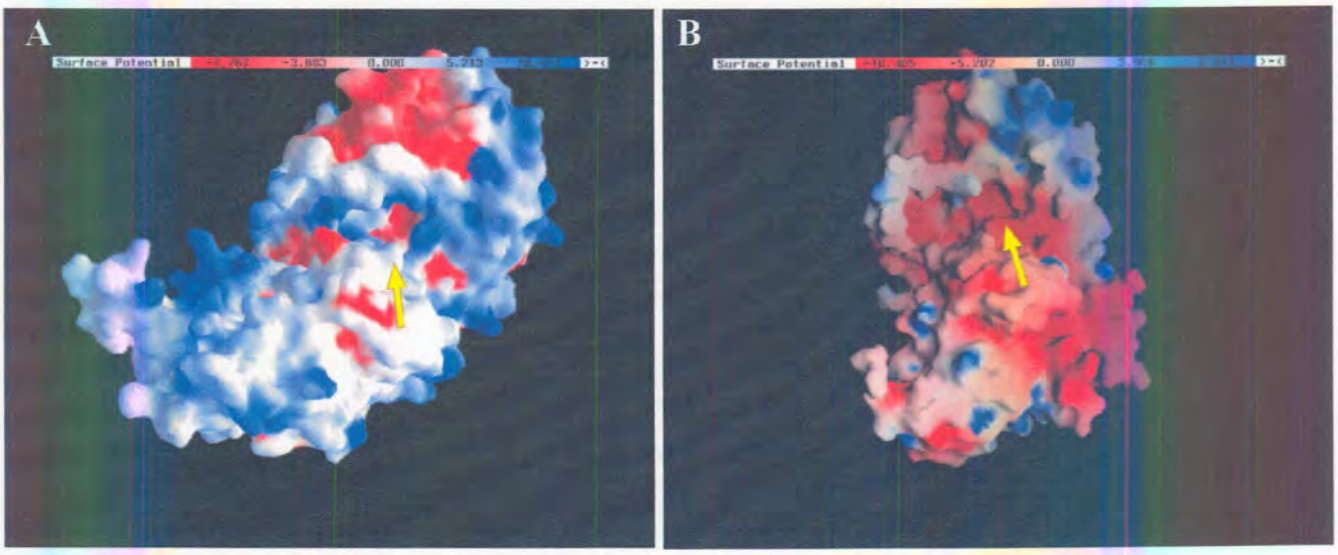

Figure 5.11: Molecular surface potentials of the monomeric PfODC (A) and human ODC (B) structures. Surfaces were created with GRASP and are coloured for electrostatic potentials with red being the most electrostatically negative and blue the most positive. Arrows indicate the view into the α/β-barrel and active site pocket.

5.3.8) Binding pocket for antizyme in PfODC.

Mammalian ODC is known to be regulated by the binding of a putrescine induced protein, antizyme. Antizyme binds to monomeric ODC in the N-terminal domain at a distinctly positively charged area called the antizyme-binding element (AzBE) and this binding is proposed to induce conformational changes in ODC to expose the C-terminal end containing a PEST region to target the degradation of ODC by the 26S proteasome (Hayashi, 1989; Hayashi and Canellakis, 1989; Hayashi and Murakami, 1995; Hayashi, *et al.*, 1996). Analyses of the corresponding area in PfODC indicate that there is only 20% sequence identity between the malarial and human ODC sequences. The α-carbon backbones of these enzymes are well conserved in this area (Fig. 5.7 B). However, the electrostatic potential at the surface of the area in PfODC corresponding to the AzBE is not as distinctly positively charged (Fig. 5.12). This is also not the case for the *T. brucei* ODC for which it was shown that antizyme does not bind the enzyme to mediate its degradation (Ghoda, *et al.*, 1990).

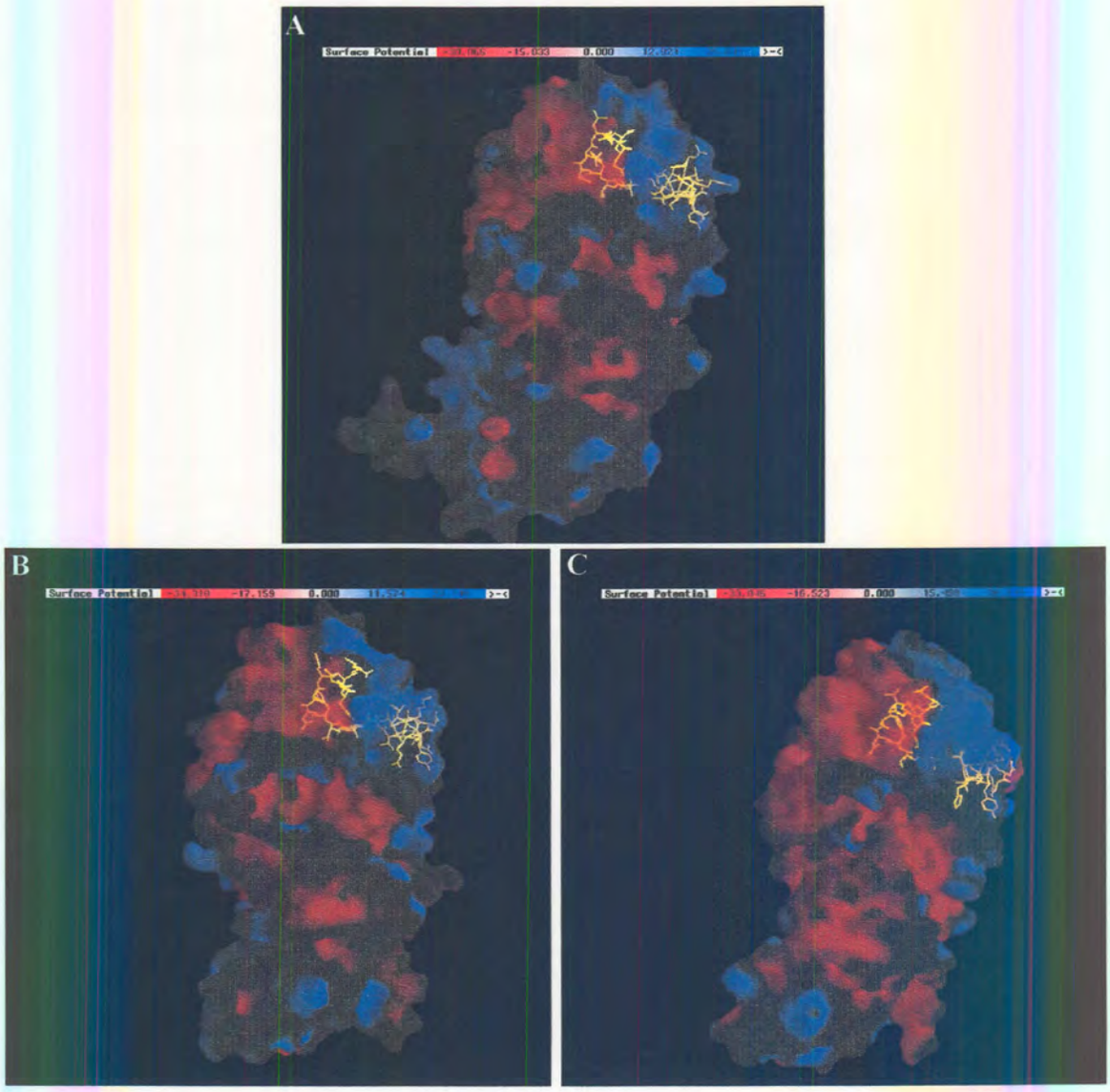

Figure 5.12: Electrostatic surface potentials for ODCs from *P. falciparum* (A), *H. sapiens* (B) and *T. brucei* (C) comparing potential antizyme binding elements. Residues forming the AzBE are indicated in yellow. Red indicates negatively charged surface areas and blue electrostatically positive areas.

5.3.9) Validation of the three-dimensional PfODC model with limited proteolysis.

Functional tests to evaluate the accuracy of the predicted PfODC model were used to delineate its predicted surface properties and domain organisation. Limited proteolysis relies on the exposure of solvent accessible areas that are selectively digested in order to reveal their organisation in the three-dimensional structure of the protein. The PfODC model was predicted to have several areas available to proteolytic splicing using the Nickpred Server. In Fig. 5.13, the model is coloured according to areas that are the most likely to be exposed to solvent and therefore be available to proteolysis (dark red areas) and areas buried in the core of the protein which is the least likely to undergo proteolysis (blue areas). This prediction shows that trypsin digestion is most likely to

occur in insert O_1 and secondly in a protease sensitive loop described in other eukaryotic ODCs (Osterman, *et al.*, 1995). Proteinase K digestion will most likely occur in the C-terminus or in insert O_1.

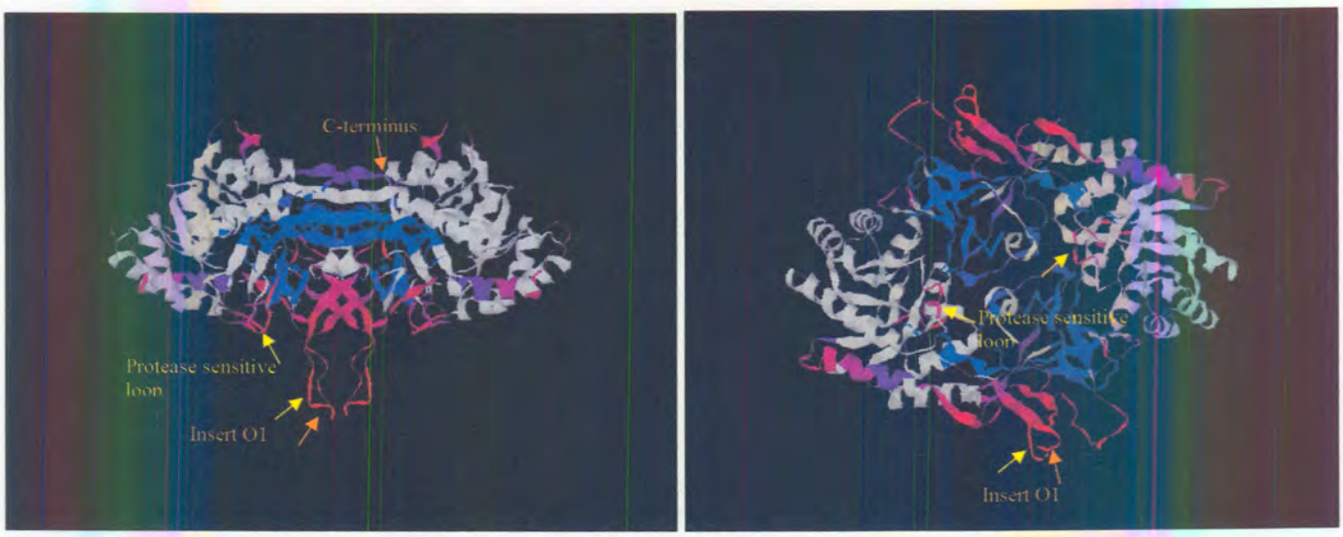

Figure 5.13: Nickpred prediction of proteolysis sites of dimeric PfODC. (A) Dimeric PfODC view from the side and in **(B)** from the top. Possible proteolytic sensitive sites are indicated in shades of red and buried areas possibly resistant to proteolysis in shades of blue. The orange arrows indicate highest scoring proteinase K prediction sites and in yellow sites for trypsin.

In order to confirm the PfODC model and the limited proteolysis predictions, the recombinantly expressed protein was subjected to diluted amounts of either trypsin or proteinase K for a short period of time. Digestion with proteinase K resulted in the wild type protein size of ~85 kDa decreasing to 82 kDa, with a 3 kDa size fragment removed (Fig. 5.14). This corresponds to the predicted cutting at the C-terminal end of PfODC where ~28 residues can be removed. Exposure of the protein to trypsin resulted in two fragments of ~45 and ~40 kDa in size. This indicates that the protein was probably cut in insert O_1. Validation of the identity of the obtained fragments was performed with peptide mass fingerprinting. However, the preliminary mass spectrometry data was inconclusive probably due to the low yield of the fragments after elution from SDS-PAGE gels.

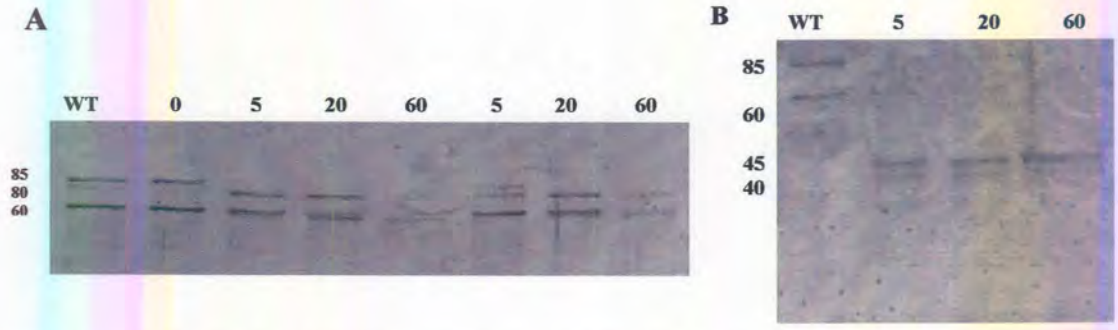

Figure 5.14: SDS-PAGE analyses of recombinantly expressed PfODC digested with either proteinase K (A) or trypsin (B). Molecular masses are indicated on the left of each figure in kDa. WT: undigested PfODC. The incubation times are indicated on top in minutes.

5.4) DISCUSSION.

The structural biology paradigm involves the determination of a protein structure to understand how the protein performs its known biological function at the molecular level (Thorton, *et al.*, 2000). Evolution has produced families of proteins whose members share the same three-dimensional architecture and frequently have detectably similar sequences. However, two structures may have very similar folds despite lacking any statistically significant sequence identity and it is now accepted that proteins having more than 30% of their sequences in common can be assumed to adopt the same folds (Jones, *et al.*, 1996). Protein structures, which are presumed to have diverged from a common ancestor in this way, are described as homologous (Jones, *et al.*, 1996). Analogous folds occur when the relationship between two structures are coincidental due to the physical limitation on protein folds. The conservation in protein folds allows a structural description of all proteins in a family even when only the structure of a single member is known (Sanchez, *et al.*, 2000). Because of the limited number of possible topologies of folds, it is a sensible approach to predict a structure by determining if a sequence could adopt one of the currently known set of protein folds (Jones, *et al.*, 1996).

5.4.1) Structural classification of PfODC.

Structural predictions are possible based on the intrinsic properties of the primary amino acid sequence of proteins. Analyses of the amino acids sequence of the PfODC with various databases and servers indicated that this decarboxylase groups into the expected family of PLP-dependent decarboxylases, with predicted α/β-barrel structure. PfODC is

grouped into the alanine racemase-like family of proteins based on sequence similarity. The proteins in this family are grouped into the TIM superfamily based on similar three-dimensional structures, which is in turn grouped into a specific protein fold, the α/β-barrel. This corresponds with the secondary structure predictions for the bifunctional PfAdoMetDC/ODC described in Chapter 3, where it was predicted that ODC contains a α-helix/β-sheet domain. All protein folds are grouped into classes of various combinations of the basic secondary structure elements of amino acids of α-helices and β-sheets. Homologous proteins sharing a known fold and having diverse functions shed light on divergent evolution. The α/β-barrel fold family is probably one of the best examples of this occurrence. The fold has been described in >20 protein families including proteins with diverse functions such as TIM, aldolase, flavocytochrome B2, tryptophan synthase, rubisco, enolase, glyoxylate oxidase and other multifold proteins including ornithine decarboxylase (Burley, 2000; Jones, *et al.*, 1996). From the above it is apparent that characteristic structural features are conserved in evolution even between proteins with diverse primary sequences.

5.4.2) Comparative modelling of PfODC.

Homologous proteins that diverged from a common ancestor and have detectable sequence similarities (30%) commonly share folds and three-dimensional architecture. This served as the underlying principle for comparative protein structural modelling (Jones, *et al.*, 1996; Srinivasan, *et al.*, 1996). The first important aspect of comparative modelling is the evaluation of the predicted models since the quality of the model determines the information that can be extracted from it. The accuracy of the obtained model can be verified by determination of the correct fold of the protein, RMSD, good stereochemistry and the distribution of spatial features (Marti-Renom, *et al.*, 2000). The results presented in this chapter indicate that the PfODC model is of a high quality that reflects the three-dimensional structure of the protein. The observed RMSD for the core of the PfODC model compared to the human ODC crystal structure compares favourably to results of the 3DCrunch project, which showed that 63% of sequences sharing 40–49% identity with their template, yield models that deviate by less than 3 Å from control X-ray crystallography structures (Guex, *et al.*, 1999). The stereochemical parameters for the α-carbon backbone indicate that the core of the PfODC model closely resembles the human ODC crystal structure indicating a 3D-structure that resembles all known eukaryotic ODC structures characterised to date. As predicted by the protein fold family analyses, PfODC contains a α/β-barrel fold. Furthermore, it also

contains a modified Greek-key β-barrel fold. However, because of its function, it is grouped into the PLP-binding superfamily of proteins that shares the α/β-barrel fold. The accuracy of the predicted PfODC structure was experimentally shown with functional tests to delineate its predicted surface properties and domain organisation. PfODC presumably undergoes limited proteolysis in either the C-terminal 28 residues or in the smaller parasite-specific insert O_1 depending on the protease used (Fig. 5.13 and 5.14). However, these results need to be confirmed with conclusive peptide mass fingerprints. The predicted surface location of the parasite-specific insert O_1 suggests its preferential digestion and not in the protease sensitive loop as is the case for the *T. brucei* ODC which should have given rise to 34 and 50 kDa bands not observed with trypsin digestion (Fig. 5.13) (Osterman, *et al.*, 1995).

5.4.3) Structural modelling of parasite-specific inserts in PfODC.

Ab initio modelling was only possible for the smallest insert O_1 in PfODC based on the intrinsic properties of the amino acids in this area and libraries of preferred side chain conformations. The insert structure is predicted to be comprised of four anti-parallel β-sheets, corresponding to the secondary structure predictions for this insert described in Chapter 4. Since no interactions could be indicated in the model between this insert and the rest of the core structure in the PfODC model, no firm conclusions are possible with regard to the exact orientation of the bulk of the loop in space. However, the insert is bridged by flexible Gly residues ($Gly_{1036-1038}$ and Gly_{1083}), which could act as hinge regions and allow mobility of the insert. The results of Chapter 4 indicated that this area is indeed important for the activity of both decarboxylases and association between the individual domains in the bifunctional PfAdoMetDC/ODC. The analyses described here predict that this area is structured and might be flexible to mediate the proposed functions. This area might therefore directly influence the active site pockets or might act as a channel in the heterotetrameric PfAdoMetDC/ODC protein and allow the substrate and co-factor to enter the hidden active site pocket in homodimeric PfODC. Substantiating evidence from the PfODC model is that this insert lies on the same face of the protein that would allow entry to the active sites. Furthermore, as mentioned in Chapter 4, this insert might influence the activity through long range energetic coupling of residues distant from the active site pocket to the catalytic residues (Myers, *et al.*, 2001).

Removal of the large parasite-specific insert (O_2) and the hinge region was necessary to create the PfODC model. The absence of 3D structural data of these inserts pre-empt conclusions on the function of these areas. The junction region between the N- and C-terminal domains of eukaryotic ODCs (region 300-340 in the murine enzyme) varies in length from 40 residues for the mouse ODC to 115 residues for the closely related *E. coli* arginine decarboxylase enzyme (Osterman, *et al.*, 1995). The insert O_2 in PfODC occurs in the equivalent region suggesting a considerable tolerance for sequence length variations in this area and a probable species-specific property. Other studies have suggested that low-complexity regions within such inserts found in malaria proteins encode for non-globular domains that occur on the surface of proteins and are not involved in the functional folding of the proteins (Pizzi and Frontali, 2001). As indicated in Chapter 4, the insert also contains a large number of Asn and Asp residues and is unstructured and nonglobular. Investigations into the conformational characteristics of asparaginyl residues indicated a peculiar feature of its side chain to consist of a peptide plane mimic attached at its C^{β} atom. It is therefore the non-glycyl residue with the most potential to adopt a left-handed α-helix conformation in the (ϕ/ψ) plane. Asn is also known to prefer loops rather than structured α-helices or β-sheets (Srinivasan, *et al.*, 1994). Modelling of these areas specifically in *P. falciparum* proteins is currently not possible.

The core structure of the PfODC described here has large similarities in its α-carbon backbone with other eukaryotic ODC structures despite the absence of the large parasite-specific insert and hinge region and the presence of parasite-specific insert O_1. Homology modelling of the malarial DHFR and serine/threonine phosphatase suggested that the inserts in these proteins occur as loops pointing away from the surface of the protein, and are proposed to be separate from the catalytic site and not to affect the models in terms of active site investigations (Lemcke, *et al.*, 1999; Li and Baker, 1998). Deletion mutagenesis of the parasite-specific inserts (Chapter 4) indicated that the large insert in the ODC domain (insert O_2) is necessary for ODC activity possibly by inducing slight conformational changes in the active site centre. It is of interest to note that single point mutations on the interface of the *T. brucei* ODC homodimer resulted in significant decreases in enzyme activity (Myers, *et al.*, 2001). However, the modelling method forces the correct conformation for the active site pocket as is evidenced by the high similarity with the human and *T. brucei* structures (See next section). It is expected

that removal of the insert would have the greatest impact in the immediate vicinity from where it was removed and that the actual structure could differ from the model.

5.4.4) Structural properties of active dimeric PfODC.

Eukaryotic ODC is an obligate homodimeric enzyme with two active sites at the interface between the monomers. PfODC is predicted to dimerise in the same manner as other eukaryotes indicating the quality of the monomeric homology models in terms of the active centres obtained. The PfODC homodimer interface is characterised by several stabilising interactions at the dimer interface. Hydrophobic interactions comparable to those in the TbODC and human ODC structures are found. More importantly, electrostatic interactions and hydrogen bonds are proposed to stabilise the associating areas between the monomers surrounding the active site pockets. Several of these interactions were unique to the PfODC model. Experimental evidence to support the predicted interactions at the PfODC dimer interface was provided by site-directed point mutagenesis of Asp_{1356}, Asp_{1359} or Lys_{970} to alanine, which led to inactive PfODC (Wrenger, *et al.*, 2001). The inability of these mutants to be catalytically active confirms the involvement of these amino acids in the dimerisation and therefore inherent activity of PfODC.

Alanine scanning mutagenesis of the *T. brucei* ODC dimer interface showed that none of the mutants caused significant weakening of the dimer interaction, suggesting that structural features contributing to dimerisation are distributed throughout the interface (Myers, *et al.*, 2001). More importantly, all the mutations caused significant detrimental effects on enzyme activity possibly by long range energetic coupling of the interface residues to the active site. It was therefore proposed that subunit interactions in ODC are optimised for catalytic function and not for high-affinity subunit association (Myers, *et al.*, 2001). Inhibitors that could dissociate the ODC dimer or bind in the interface and disrupt activity would have advantages over traditional active site-directed inhibitors as mentioned in Chapter 4. Since there are discrete regions predicted to be involved at the dimer interface in PfODC, this could have implications in the selective inhibition of ODC activity in the malaria parasite. However, as this activity is part of a bifunctional protein, it remains to be seen if effects predicted for the homodimeric monofunctional PfODC can be extrapolated to this protein in complex with PfAdoMetDC.

The homodimeric model of PfODC reveals two identical active sites formed at the dimer interface by contributions from both monomers. There is a large degree of residue conservation and similar spatial orientations between the active site pockets of PfODC and the ODCs from *T. brucei* and *H. sapiens*. Single substitution of conserved residues Lys_{868}, Cys_{1355} or Asp_{1356} with alanine abolished PfODC enzyme activity, providing experimental support for the role of these residues in the activity of PfODC through the predicted interactions with the substrate or co-factor as seen in the model and lends support to the accuracy of the predicted active site pocket in the model (Wrenger, *et al.*, 2001). However, parasite-specific residues (Thr_{933}, Asn_{1034}, Met_{967}, Arg_{1117} and Tyr_{966}) were identified in the PfODC model and predicted to be selectively involved in stabilising interactions with the co-factor and substrate. This allows further mapping of the PfODC active site pocket by point mutations to confirm its accuracy. Once this is confirmed, the design of parasite-specific inhibitors could be considered. Furthermore, the electrostatic potential of the PfODC active site pocket is significantly different to the corresponding human ODC. This could have positive implications to the selective entry of PfODC-specific inhibitors into the active site pocket of PfODC.

5.4.5) Potential role of antizyme in the regulation of PfODC based on structural properties.

Eukaryotic ODC is one of the most highly regulated enzymes described to date. It has an extremely short half-life and its degradation has been shown to be mediated by an inhibitory protein, antizyme. Antizyme binding not only inactivates ODC by preventing dimerisation but also induces conformational changes to expose a basal-degradation element in the C-terminus of the protein and targets it for degradation by the 26S proteasome in an ubiquitin independent manner (Almrud, *et al.*, 2000; Hayashi and Canellakis, 1989; Hayashi and Murakami, 1995; Hayashi, *et al.*, 1996; Heller, *et al.*, 1976). A distinct binding site for antizyme could not be shown in the PfODC model. Antizyme might therefore not be able to bind PfODC and regulate it in the same manner as in other organisms. The limited proteolysis of PfODC indicated that the C-terminus is exposed to solvent. However, unlike the human ODC, the PfODC does not contain a significant PEST region in this area (Chapter 3, Fig. 3.12)(Almrud, *et al.*, 2000). ODC usually has a rather long half-life in parasitic protozoa and antizyme mediated degradation of ODC has not been described for any of the parasitic protozoa as mentioned in Chapter 3.

Chapter 6 describes how the homology model for PfODC presented in this chapter is used to explain the experimental inhibition of the ODC component of PfAdoMetDC/ODC with various inhibitors and explores the rational design of novel ODC inhibitors as possible lead compounds for antimalarial chemotherapy.