Chapter 3

Potential Quaternary Structure of *Plasmodium S*-Adenosylmethionine Decarboxylase/Ornithine Decarboxylase

3.1 Introduction

3.1.1 Ornithine decarboxylase

3.1.1.1 Structure and reaction mechanism

Ornithine decarboxylase (ODC) catalyses the decarboxylation of ornithine to release putrescine (EC 4.1.1.17, Fig. 3.1) and requires the co-factor pyridoxal 5'-phosphate (PLP). Two structural classes of PLP-dependent decarboxylases have been identified. The first family is homologous with the aspartate aminotransferase-like superfamily (Sandmeier *et al.*, 1994) and largely found in prokaryotes. The second class is homologous with the alanine racemase superfamily, and comprises an *N*-terminal TIM barrel domain and a C-terminal β-sheet domain (Grishin *et al.*, 1995) of about 45 kDa. The second class includes arginine decarboxylase, diaminopimelate decarboxylase and recently lysine/ornithine decarboxylase in prokaryotes (Takatsuka *et al.*, 2000; Lee *et al.*, 2007). *P. falciparum* ODC is a member of the second class of PLP-decarboxylases.

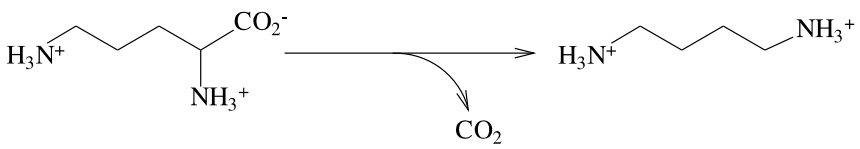


Figure 3.1: Ornithine decarboxylase reaction

The first ODC structure of the second class was determined for the mouse enzyme (Kern *et al.*, 1999). Since then structures have also been determined from humans, *Trypanosoma*

brucei and *Vibrio vulnificus* (lysine/ornithine decarboxylase) (Grishin *et al.*, 1999; Almrud *et al.*, 2000; Lee *et al.*, 2007). Unlike most TIM barrels which begin with a $\beta-$strand, the barrel domain begins with helix 2 and ends with helix 10 (Fig. 3.2). The sheet domain is largely made up of C-terminal residues and comprises two sheets which twist into each other. Sheet one consists of seven strands ($\beta 2 \downarrow -\beta 18 \downarrow -\beta 3 \downarrow -\beta 12 \uparrow -\beta 17 \downarrow -\beta 14 \uparrow$) and is connected by $\beta 12$ to sheet two which consists of four strands ($\beta 12 \uparrow -\beta 13 \downarrow -\beta 15 \downarrow -\beta 16 \uparrow$).

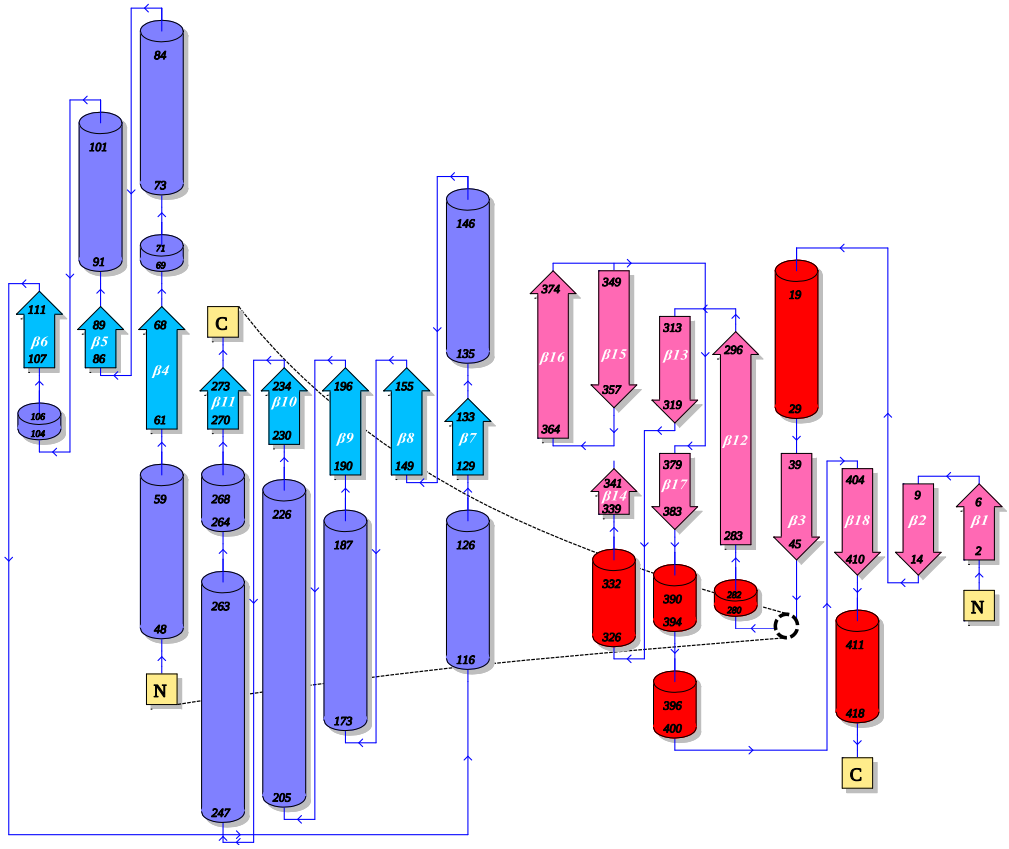


Figure 3.2: Topology of ODC TIM barrel (left) and sheet (right) domains, PDB id: 7ODC. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The insertion site of the TIM barrel within the sheet domain is indicated by the dashed circle. The figure was adapted from output generated on the PDBsum server (http://www.ebi.ac.uk/pdbsum/).

The active site with PLP is located at the dimer interface with residues donated from both monomers, making it an obligate dimer. PLP is covalently bound to Lys 69 (Mouse/*T. brucei* ODC) as an internal aldimine (Osterman *et al.*, 1999). During the reaction the lysine residue is displaced by the substrate via a *gem*-diamine intermediate (Jackson *et al.*, 2004) to form the external aldimine (Fig. 3.3). In PLP co-factor enzymes, reactions differ from this point depending on which of the C_α bonds (except $C_\alpha - N$) is cleaved. This leads to a wide variety of possible reactions including decarboxylation, transamination, racemisation, β-elimination and retro-aldol cleavage. Reaction specificity is thus determined by alignment of the bond to be cleaved perpendicular with the plane of the PLP ring (Toney, 2005). In the mechanism

proposed by Dunathan (1966) alignment of the p-orbital with the conjugated π-system of the PLP ring promotes labilisation of the target bond. Specificity is thus determined by active site binding of substrate moieties. This mechanism has been questioned recently by quantum chemistry studies of proposed PLP-substrate transition states, as well as the existence of pyruvoyl-based decarboxylases that function without the PLP-ring. Instead, it has been suggested that the PLP ring serves to stabilise the carbanionic intermediate (Bach *et al.*, 1999; Toney, 2001). Following elimination of CO_2 the C_α is reprotonated and the substrate exchanges with Lys 69 to reform the internal aldimine (Fig. 3.3). The acid for re-protonation is mostly likely Cys 360′ (human). Cys 360′ rotates into the active site upon substrate binding to within 3.4 Å of the imine bond, and mutation to Ser or Ala reduces k_{cat} substantially and promotes the side-reaction of re-protonation at $C4'$ instead (Grishin *et al.*, 1999; Jackson *et al.*, 2000). The carboxyl group is oriented on the *re* face of PLP in a hydrophobic pocket that is thought to favour cleavage of the leaving carboxyl to accommodate neutral CO_2. Mutation of Phe 397 to Ala within this pocket decreases the rate of de-carboxylation 2100× fold (Jackson *et al.*, 2003a). A number of other residues have been established to be important for substrate binding. Glu 274 interacts with the PLP nitrogen (Grishin *et al.*, 1999) while the phosphate binds to a highly conserved glycine-rich loop (Kern *et al.*, 1999). Substrate specificity for ornithine is determined by interactions between the amino group of ornithine and Asp 332 and Asp 361′ (Grishin *et al.*, 1999) (Fig. 3.4).

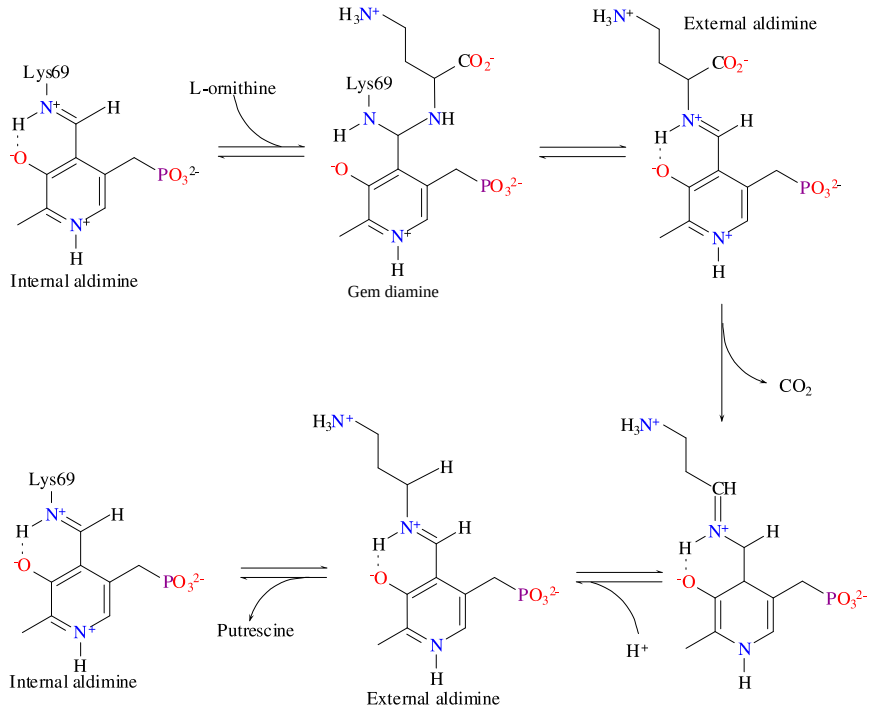


Figure 3.3: Reaction mechanism of ornithine decarboxylase, adapted from Jackson *et al.* (2003a). The incoming substrate displaces Lys 69 to form the external aldimine via a gem diamine intermediate. The reaction that follows depends on the orientation of the moieties around C_α relative to the PLP plane. In ODC the $C_\alpha - CO_2$ is perpendicular to PLP. CO_2 is then released followed by reformation of the internal aldimine. For further details see text.

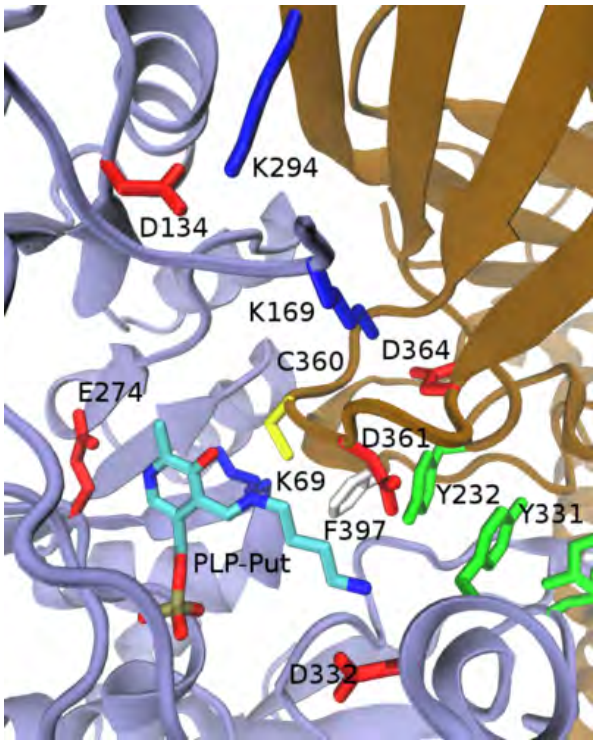


Figure 3.4: Active site and dimer interface residues in ornithine decarboxylase from *T. brucei* with the PLP-Putrescine (PLP-Put) adduct. Both ODC monomers (blue and brown, respectively) contribute residues to the active site. Acidic (red), basic (blue), cystien (yellow) residues are indicated The specificity of PLP cofactor enzymes is determined by the orientation of $C_\alpha - R$ moieties relative to the PLP ring. See text for further details.

3.1.1.2 Quaternary structure

The two monomers of ODC associate in a symmetrical head-to-tail fashion (Fig. 3.5). The barrel domain of one monomer contacts the sheet domain of the other. Considerable variation is observed in the surface area of the dimer interface, ranging from 655 Å^2 in mouse ODC to 2775 Å^2 in *T. brucei* ODC(Kern *et al.*, 1999; Grishin *et al.*, 1999). A number of conserved dimer interactions have been identified that are necessary for enzyme activity. These include salt-bridges between Lys 169 and Asp 364′ as well as Asp 134 and Lys 294′. Additionally, a hydrophobic zipper is formed by Phe 397, Tyr 323 and Tyr 331′ (Fig. 3.4). Mammalian ODC is known to undergo rapid subunit exchange (Coleman *et al.*, 1994), and this has been attributed in part to the small dimer interface. Rapid exchange has also been observed for the *T. brucei* enzyme, however, which can also form cross-species heterodimers with mouse ODC (Osterman *et al.*, 1994). Mammalian ODC undergoes rapid turnover via ATP-dependent non-ubiquitin mediated proteolysis. This pathway is mediated by the inhibitor protein antizyme, which binds to the *N*-terminal barrel domain and targets the inactivated monomer for the proteosome (Coffino, 2001). Antizyme is itself inhibited by antizyme inhibitor, a close enzymatically inactive homologue of ODC (Albeck *et al.*, 2008).

ODC from *T. brucei* is also observed to be susceptible to mutations and ligand binding distant from the active site. Myers *et al.* (2001) demonstrated 11 mutations more than 10 Å from the active site that substantially decreased enzyme activity. Of these, Lys 294 Ala was later observed to cause disorder in a loop (Leu 166 - Ala 172) involved in maintaining the dimer interface (Jackson *et al.*, 2004). Furthermore, the non-competitive inhibitor G418 (Geneticin) binds allosterically at the dimer interface and also induces disorder in an active site loop (Val 392 - Gln 401) (Jackson *et al.*, 2003b). These results suggest that the ODC

Figure 3.5: Quaternary structure of ornithine decarboxylase (*T. brucei*. Monomer A (Increasing residue number: blue-green-red), monomer B (purple).

class of enzymes may be susceptible to non-active site targeting.

3.1.2 *S*-Adenosylmethionine Decarboxylase

3.1.2.1 Structure and reaction mechanism

S-Adenosylmethionine decarboxylase (AdoMetDC) catalyses the decarboxylation of *S*-adenosylmethionine to 5′-(3-aminopropylmethylsulphonio)-5′-deoxyadenosine (EC: 4.1.1.50, Fig. 3.6). AdoMetDC comprises two subunits which are the result of an internal proteolytic cleavage at the active site. This generates the smaller *N*-terminal β-subunit and the larger *C*-terminal α-subunit. At the cleavage site a pyruvoyl residue is generated from the *N*-terminus of the α-subunit which functions in a similar manner to PLP and is essential for enzyme activity. The AdoMetDC fold was thought to be singular (Kozbial and Mushegian, 2005). However, a novel kind of arginine decarboxylase was recently discovered in the hyper-thermophilic *Crenarchaeota*, *Sulfolobus solfataricus*. Within this class of organisms AdoMetDC and arginine decarboxylase are thought to be paralogues that diverged from a common ancestor after gene duplication (Giles and Graham, 2008).

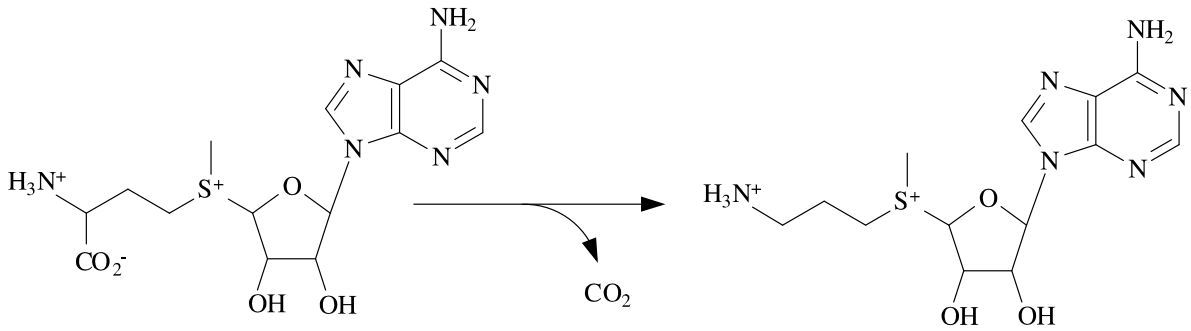


Figure 3.6: *S*-Adenosylmethionine decarboxylase reaction.

The eukaryotic AdoMetDC fold comprises an $\alpha\beta\beta\alpha$ sandwich of which AdoMetDC is the

most studied member. The active site is located between the two β-sheets at the proteolytic cleavage site. To date structures have obtained for human and potato (*Solanum tuberosum*) AdoMetDC (Ekstrom *et al.*, 1999; Bennett *et al.*, 2002) and from the bacteria *Thermatoga maritima* (Toms *et al.*, 2004). An unpublished structure for *Aquifex aeolicus* (PDB id: 2III) has also been deposited in the PDB. The two $\alpha\beta$-halves of eukaryotic AdoMetDC are topologically equivalent. The bacterial structure is a dimer of $\alpha\beta$ subunits with two active sites which associate via stacking of the β-sheets. Each of these subunits are in turn topologically similar to each $\alpha\beta$-half of eukaryotic AdoMetDC. Eukaryotic AdoMetDC is thus thought to have resulted from an internal duplication of the bacterial gene, with subsequent loss of the *C*-terminal active site (Toms *et al.*, 2004). In eukaryotic AdoMetDC each β-sheet comprises eight strands with the following topology: $\beta 2 \uparrow -\beta 3 \downarrow -\beta 4 \uparrow -\beta 5 \downarrow -\beta 1 \uparrow -\beta 6 \downarrow -\beta 8 \uparrow -\beta 7 \downarrow$ (Fig. 3.7). *Thermotoga maritima* AdoMetDC is an $(\alpha\beta)_2$ dimer, each $\alpha\beta$ portion containing an anti-parallel β-sheet of six strands with the following topology: $\beta 2 \uparrow -\beta 3 \downarrow -\beta 4 \uparrow -\beta 5 \downarrow -\beta 1 \uparrow -\beta 6 \downarrow$ (Fig. 3.8).

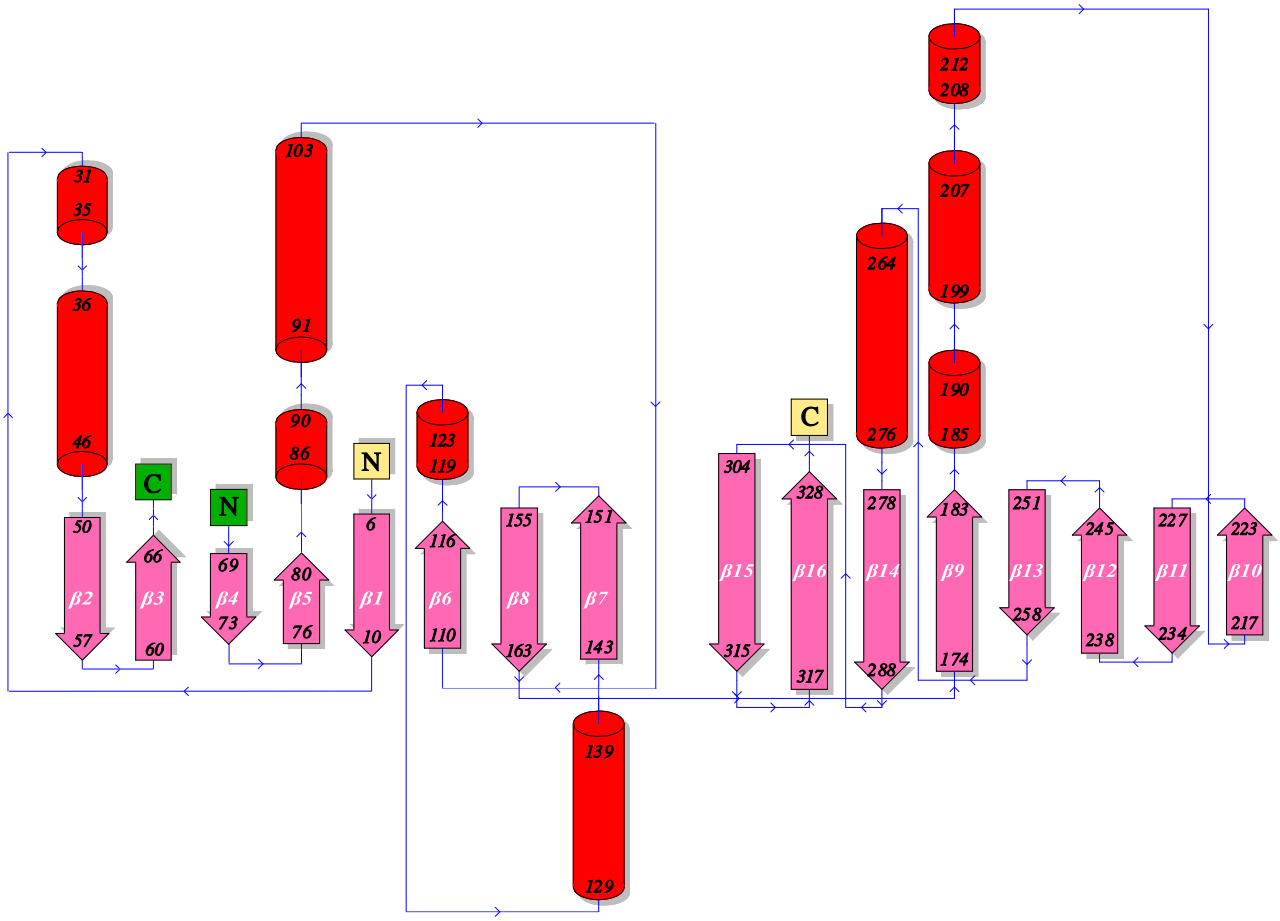


Figure 3.7: Topology of eukaryotic (human) AdoMetDC, PDB id: 1JEN. The site of internal proteolytic cleavage is indicated in green. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The figure was adapted from topology diagrams generated on the PDBsum server (http://www.ebi.ac.uk/pdbsum/) for the β- and α-chain.

The generation of the active site pyruvoyl is a non-hydrolytic serinolysis reaction (Fig.

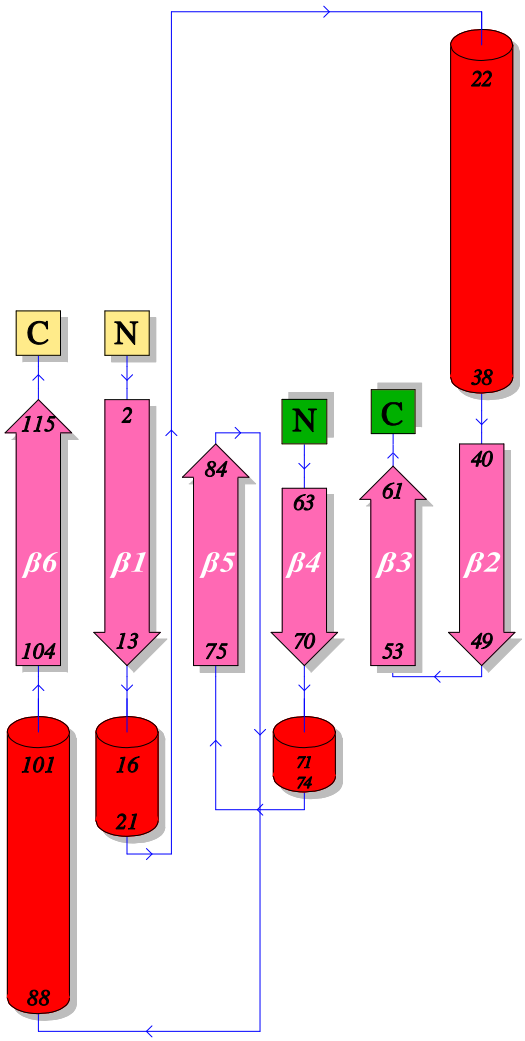


Figure 3.8: Topology of bacterial AdoMetDC from *Thermotoga maritima*, PDB id: 1TLU. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The figure was adapted from that generated on the PDBsum server (http://www.ebi.ac.uk/pdbsum/) for the α- and β-chain. The site of internal proteolytic cleavage is indicated in green.

3.9). In human AdoMetDC, Ser 68 donates its side-chain oxygen to the *C*-terminus and the remainder of the residue converts to ammonia and the pyruvoyl group. The reaction proceeds via an $N \rightarrow O$ acyl rearrangement (Recsei and Snell, 1984; van Poelje and Snell, 1990) via an oxyoxazolidine intermediate. This is followed by cleavage of the ester by β-elimination to yield a normal carboxyl terminus and dehydroalanine on the *N*-terminus. The dehydroalanine tautomerises to an imine which is converted to a carbinolamine. Deamination yields the pyruvoyl group (Tolbert *et al.*, 2003). The His 243 Ala mutant remains in the ester intermediate, suggesting that His 243 is the base required for extraction of the H_α proton during β-elimination. Glu 11 may serve to stabilise His 243, thus potentially forming a catalytic triad with Ser 68 (Ekstrom *et al.*, 2001) typically found in proteases. Cys 82 may assist processing by protonation of the carbonyl oxygen of Glu 67 during nucleophillic attack by Ser 68 on the carbonyl group. Mutating Cys 82 to Ala or Ser slows down processing approximately 10-fold (Stanley and Pegg, 1991; Tolbert *et al.*, 2003). Ser 229 Ala results in a non-processing mutant while Ser 229 Thr processes readily. Ser 229 is thought to stabilise the oxyoxazolidine intermediate through hydrogen bond interactions (Xiong and Pegg, 1999; Tolbert *et al.*, 2003).

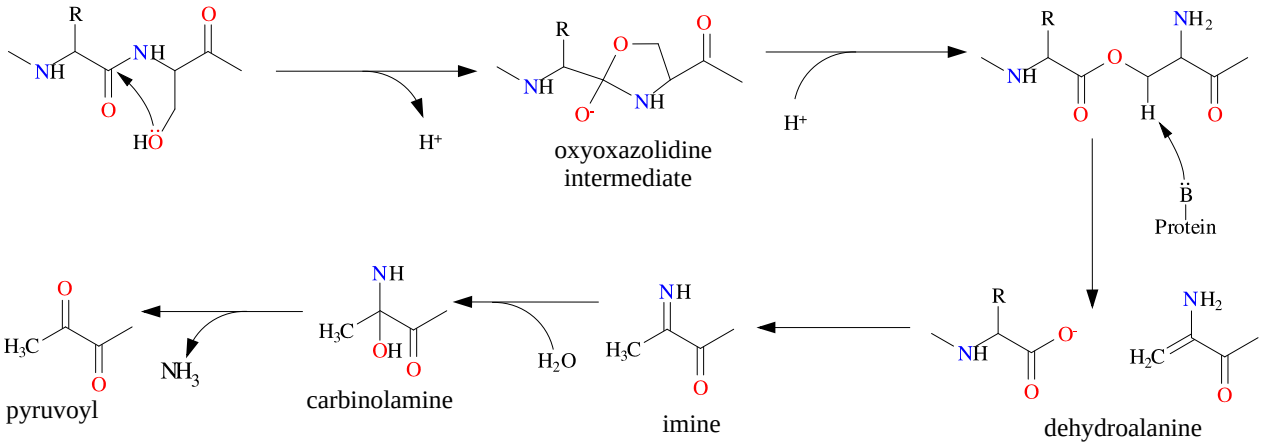


Figure 3.9: Pyruvoyl generation mechanism of *S*-adenosylmethionine decarboxylase takes place via an internal non-hydrolytic serinolysis. See text for further details. Adapted from Tolbert *et al.* (2003).

The generally accepted reaction mechanism of AdoMetDC is depicted in Figure 3.10. It is similar to PLP decarboxylases in that the pyruvoyl residue forms a Schiff base with the amino acid moiety. This facilitates cleavage of the α-carboxylate group and release of CO_2. Re-protonation of the $C_\alpha$atom allows hydrolysis of the Schiff base and product release. AdoMetDC is unusual in that the *S*-adenosylmethionine binds with the adenosyl ring in the *syn*- conformation rather than the usual *anti*- conformation. The *anti*- conformation is facilitated by hydrophobic stacking of the adenosyl ring between the side-chains of Phe 7 and Phe 223 (human AdoMetDC). The ribosyl group is stabilised by hydrogen bond interactions with Glu 247. Ser 229, His 243 and Cys 82 known from mutagenesis studies to be important for activity are all located near to the scissile bond and pyruvoyl group. Cys 82 and His 243 are both candidates for protonating the decarboxylated Schiff base. Cys 82 is the most likely proton donor with His 243 possibly functioning in this capacity at a lower capacity. Ser 229 is within hydrogen bonding distance of His 243 and may help prevent improper protonation (Tolbert *et al.*, 2001). The normal substrate of AdoMetDC carries a positive change in the form of the sulphonium group and substrate analogues which lack a similar positive charge don't display significant inhibition (Pankaskie and Abdel-Monem, 1980; Pegg and Jacobs, 1983). No negatively charged residue has been identified to interact with this group. However, it has recently been determined by quantum chemical calculations that cation-π interaction between Phe 7, Phe 223 and the sulphonium atom are responsible for determining this specificity (Bale *et al.*, 2009).

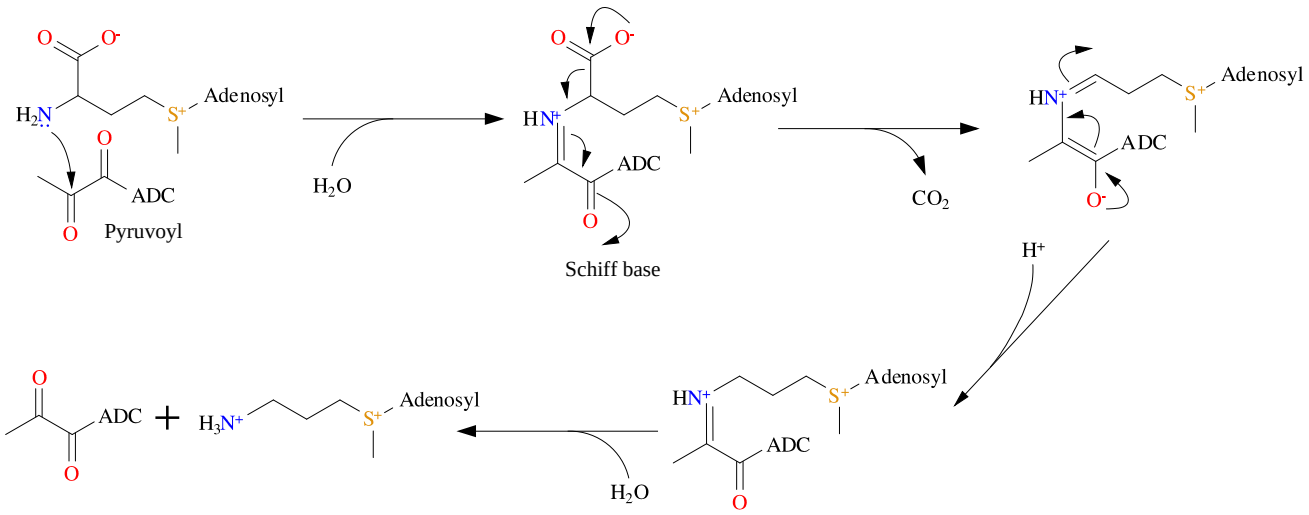


Figure 3.10: Reaction mechanism of *S*-adenosylmethionine decarboxylase. The pyruvoyl residues acts in much the same way as PLP-based decarboxylases. Formation of a Schiff base facilitates cleavage of the α-carboxylate group and CO_2. See text for further details. Adapted from Tolbert *et al.* (2001).

3.1.2.2 Quaternary structure and allosteric regulation

Eukaryotic AdoMetDC is found either as a monomer (one $\alpha-$ and one $\beta-$subunit each) or dimer depending on the species. AdoMetDC from humans and Trypanosoma is observed to dimerise (Ekstrom *et al.*, 1999; Clyne *et al.*, 2002), while that from potato is monomeric (Bennett *et al.*, 2002). Dimerisation in human AdoMetDC occurs via an edge-on association between the β-sheets, β-strand 15 and β-strand 7 that interact with their counterparts in the adjacent monomer via main-chain and side-chain hydrogen bonds. The active sites in the human AdoMetDC dimer are 54 Å apart and far removed from the dimer interface (Ekstrom *et al.*, 1999). Bacterial AdoMetDC either forms an $(\alpha\beta)_4$ tetramer in Gram-negative species (Markham *et al.*, 1982) which corresponds to the eukaryotic dimer, or a an $(\alpha\beta)_2$ dimer in Gram-positive and Archeabacteria (Sekowska *et al.*, 2000; Kim *et al.*, 2000), akin to the eukaryotic monomer.

In a number of species AdoMetDC is activated by putrescine, including mammals (Kameji and Pegg, 1987; Stanley *et al.*, 1994), nematodes (Da'Dara *et al.*, 1996; Da'dara and Walter, 1998), fungi (Hoyt *et al.*, 2000) and *Trypanosoma* (Kinch and Phillips, 2000). In humans putrescine increases both the rate of processing and activity. The putrescine binding site in human AdoMetDC is situated approximately 15 Å from the active site between the β-sheets (Fig. 3.11). One pyruvoyl binds per monomer. The ammonium groups of putrescine form polar and electrostatic interactions with Glu 15, Asp 174, Thr 176, Glu 178, Glu 15 and Glu 256. The aliphatic chain of putrescine stacks against Phe 285 and Phe 111. The binding of putrescine is thought to activate the enzyme through electrostatic effects affecting the orientation of key active site residues (Bale *et al.*, 2008). Plant AdoMetDC does not display any response to putrescine and is constituitively active (Xiong *et al.*, 1997). In the corresponding site in potato AdoMetDC a number of key substitutions provide moieties

that mimic putrescine (Bennett *et al.*, 2002). Putrescine has not been observed to affect bacterial AdoMetDC, however divalent cations such as Mg^{2+} are required by tetrameric AdoMetDC for activation (Markham *et al.*, 1982). In *Trypanosoma cruzi* putrescine is observed to stimulate activity but not processing (Clyne *et al.*, 2002; Beswick *et al.*, 2006). Trypanosamal AdoMetDC dimerises, however a more important regulatory mechanism was recently discovered. Trypanosomal AdoMetDC from both *T. brucei* and *T. cruzi* is observed to form a heterodimer with prozyme, an inactive non-processing homologue of AdoMetDC unique to Trypanosomatids. While the effect of putrescine is modest ($<$ 10 fold change in k_{cat}) compared to mammalian AdoMetDC, the binding of prozyme is observed to induce a 100-1000 fold change in k_{cat} in the absence of putrescine (Willert *et al.*, 2007; Willert and Phillips, 2009).

(a) (b)

Figure 3.11: Quaternary structures of AdoMetDC. (a) Eukaryotic (human) AdoMetDC dimer. In the left monomer the α-chain (grey) and β-chain (black) is indicated. The ligands MeAdoMet (red spheres) and putrescine (magenta spheres) are indicated in the right monomer. The MeAdoMet binding site is also the site of proteolytic cleavage. Sequence position is indicated in the right monomer by a blue→green→red colour gradient (b). Prokaryotic (*T. maritima*) AdoMetDC dimer (which corresponds to one eukaryotic monomer). The site of proteolytic cleavage at Ser 63 is indicated (cyan sphere). Increasing sequence position is indicated in the monomer by a blue→green→red colour gradient (See text for further details).

3.1.3 Bifunctional *Plasmodium* AdoMetDC/ODC

In most species studied to date AdoMetDC and ODC functionalities occur in separately expressed proteins. However, in the malaria parasite *P. falciparum*, both functionalities were found in a singly expressed protein (Müller *et al.*, 2000). AdoMetDC forms the *N*-terminal domain, connected by a hinge region to the *C*-terminal ODC domain. Two AdoMetDC/ODC bifunctional proteins associate to form a dimeric complex [1]. Each AdoMetDC/ODC is about

[1] In the literature the complex is sometimes referred to as being heterotetrameric (Müller *et al.*, 2000), by virtue of the generation of separate polypeptides after auto-proteolytic cleavage in the AdoMetDC active

170 kDa, while the entire complex is about 330 kDa. Cleavage generates the β-fragments of about 9 kDa each.

Unlike it's mammalian counterpart, *P. falciparum* ODC is inhibited by putrescine (Krause *et al.*, 2000). Furthermore, putrescine does not stimulate AdoMetDC (Wrenger *et al.*, 2001) as is observed for mammals or *Trypanosoma*. This was rationalised from homology modeling by the presence of internal residues that mimic the function of putrescine (Wells *et al.*, 2006), as was first discovered for plant AdoMetDC (Bennett *et al.*, 2002).

The entire AdoMetDC/ODC sequence comprises 1419 amino acids, making it considerably larger than the combined length of typical AdoMetDC ($\pm$ 320 residues) and ODC ($\pm$ 400 residues) from other organisms. Despite the extra size contributed by the hinge region ($\pm$ 300 residues), the *P. falciparum* AdoMetDC ($\pm$ 520 residues) and ODC ($\pm$ 610 residues) domains are considerably longer. The extra size is due to the presence of parasite-specific inserts within the core AdoMetDC and ODC domains compared to homologues. Initially, three inserts were identified, one in AdoMetDC and two in ODC (O_1 and O_2, Fig. 3.12 and Fig. 3.13). During subsequent studies to predict the structures of ODC and AdoMetDC via homology modeling (Birkholtz *et al.*, 2003; Wells *et al.*, 2006) the number of inserts in AdoMetDC was refined to three inserts: A_1, A_2 and A_3 (Fig. 3.12 and Fig. 3.14). *Plasmodium* proteins are unusual in that they often contain large inserts relative to their homologues (Gardner *et al.*, 1998, 2002). These inserts are generally characterised by being rich in hydrophilic residues and often of low complexity. They are predicted to be non-globular structures that protrude from the core fold (Pizzi and Frontali, 2000, 2001). While the evolutionary significance for these inserts is still being debated, the inserts specific to AdoMetDC/ODC have been found to affect the functioning of the bifunctional enzyme (Birkholtz *et al.*, 2004). This is despite the apparent lack of regulation between the active sites of AdoMetDC and ODC within the bifunctional complex (Wrenger *et al.*, 2001). Deletion of inserts O_1 and O_2 reduces activity not only in ODC but the AdoMetDC domain as well. Similar results are obtained for deletion of a region that contains A_3, however this deletion contains regions that were later predicted to form part of the core AdoMetDC structure (Wells *et al.*, 2006). The O_1 insert also appears to be involved in the formation of the complex between AdoMetDC and ODC when the domains are separately expressed (Birkholtz *et al.*, 2004).

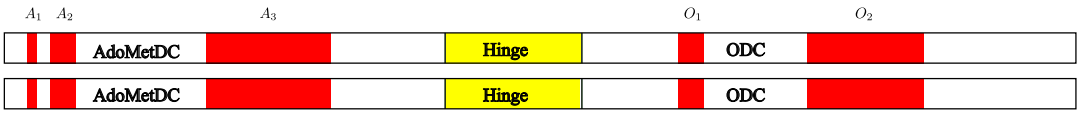


Figure 3.12: Linear structure of the *P. falciparum* AdoMetDC/ODC bifunctional complex. The domains, parasite-specific inserts (red) and hinge (yellow) regions are indicated.

Since the discovery of bifunctional AdoMetDC/ODC in *P. falciparum* the sequences for the bifunctional cognates from other *Plasmodium sp.* have since been discovered by the *Plasmodium* genome sequencing project (Gardner *et al.*, 2002). Currently complete

site. In this work a functional view at domain level is adopted, hence the complex is referred to as a dimer.

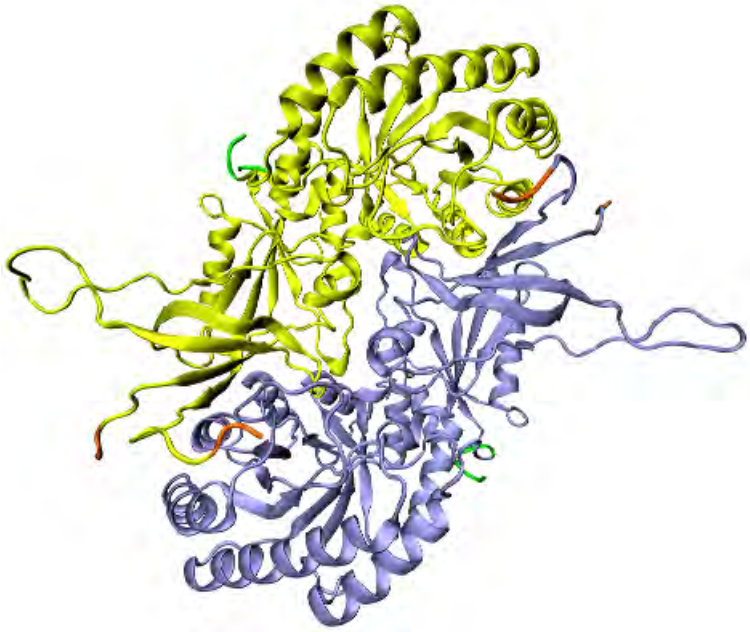


Figure 3.13: Homology model of *P. falciparum* ODC. Yellow, blue: monomers. Parasite-specific inserts are indicated in green (O_1) and orange (O_2). These inserts were too long to be modelled *ab initio* and most residues were thus excluded.

Figure 3.14: Homology model of *P. falciparum* AdoMetDC. Yellow: α-chain, blue: α-chain. Parasite-specific inserts are indicated in green (A_1) , orange (A_2) and red (A_3). Inserts and were too long to be modelled *ab initio* and thus most residues were thus excluded.

sequences can be obtained for *P. berghei*, *P. knowlesi*, *P. vivax* and *P. yoelii* from the online *Plasmodium* genome resource PlasmoDB (Bahl *et al.*, 2003). The shorter inserts of AdoMetDC and ODC are more conserved in length and composition across *Plasmodium* species (Birkholtz *et al.*, 2004; Wells *et al.*, 2006). In AdoMetDC, A_1 is seven residues long while A_2 is 26-27 residues long, while in ODC O_1 is 38 residues long. In contrast A_3 shows considerable variability in length (150-250 residues) and sequence composition between *Plasmodium sp.* Insert O_2 varies in length between 120-130 residues and also shows considerable sequence variation.

3.2 Aims

The unique bifunctional organisation of *Plasmodium* AdoMetDC/ODC is considerably different compared to human host AdoMetDC and ODC. Specifically the occurrence of these functions within one protein complex and the cross-domain and insert effects suggest that these enzymes can be interfered with by non-active site inhibitors. In order to explore this possibility, further knowledge of the most likely interacting surfaces between AdoMetDC and ODC is required. This can be used to prioritise regions to be targeted via methods such is site-directed mutagenesis, peptide mimics, etc. Because of the lack of structural data for these enzymes, *in silico* protein-protein docking and other computational analysis of AdoMetDC and ODC models was performed to determine the most likely contact regions.

3.3 Methods

3.3.1 Docking of AdoMetDC and ODC

3.3.1.1 Modeling of AdoMetDC/ODC

Both AdoMetDC and ODC domains of *P. falciparum* had been previously modelled in separate exercises (Birkholtz *et al.*, 2003; Wells *et al.*, 2006). However, since those efforts some new techniques were learnt and newer programs introduced. Both domains were therefore re-modelled using the same newer methodologies and software. Furthermore, models were also generated for all *Plasmodium* species for which a complete sequence of the bifunctional protein had been identified from genome sequencing projects. The species modelled were: *P. falciparum*, *P. berghei*, *P. knowlesi*, *P. vivax* and *P. yoelii*. For the new round of modeling a few changes to the alignments were made (Fig. 3.15 - 3.17).

For modeling AdoMetDC the human (PDB id: 1I7B, Tolbert *et al.* 2001) and potato (*Solanum tuberosum*, PDB id: 1MHM, Bennett *et al.* 2002) templates were used. MODELLER9v2 was used for homology modeling. A restraint was used to keep Pro 164 of *P. falciparum* and its cognates in other *Plasmodium* sequences (human: Pro 126, potato: Pro 129) in the *cis* conformation of backbone angles. A slow refinement was used to generate 100 models with different randomisations. The cluster method was also used to generate an average of the

100 models.

For modeling ODC the human (PDB id: 1D7K Almrud *et al.* 2000) and *Trypanosoma brucei* (PDB id: 1F3T Jackson *et al.* 2000) templates were used. The full dimer was modelled in order to provide a more realistic surface to dock the smaller AdoMetDC models against. Restraints were added to all atoms of the monomer in order to build a symmetrical dimer. Multiple models were constructed as for AdoMetDC.

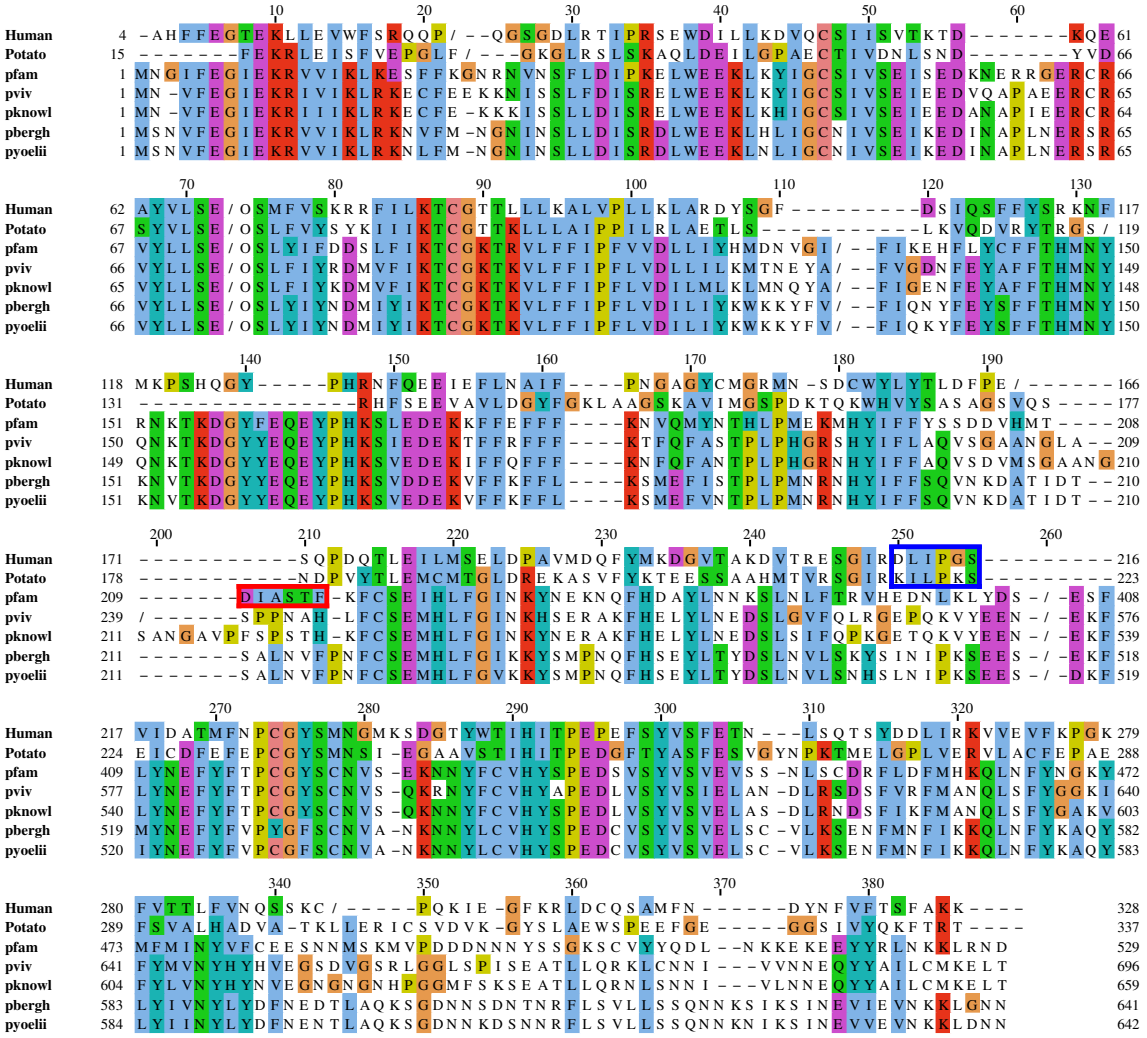


Figure 3.15: Alignment for modeling AdoMetDC. Human and *S. tuberosum* (potato) were used as templates, *P. berghei* (bergh) *P. falciparum* (pfam) *P. knowlesi* (pknowl) *P. vivax* (pviv) *P. yoelii* (pyoelii). The groups with red and blue borders were initially aligned one residue upstream (Wells *et al.*, 2006). The group in red was adjusted to account for a deleted proline residue in the primate-infecting *Plasmodium* sequences. The group in blue was adjusted one residue upstream to give better alignment with the other *Plasmodium* sequences, and also removed an insertion within an α-helix.

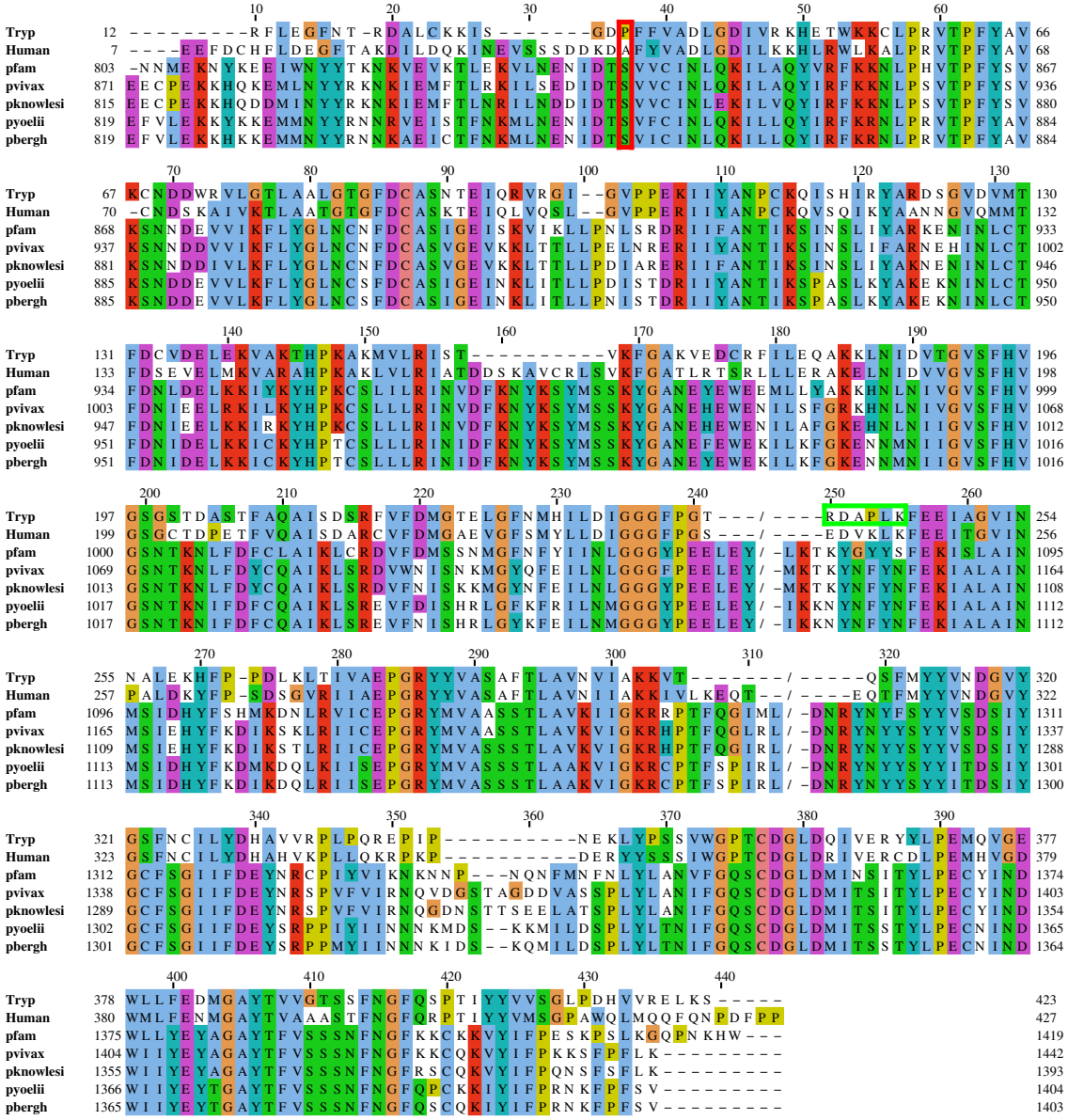
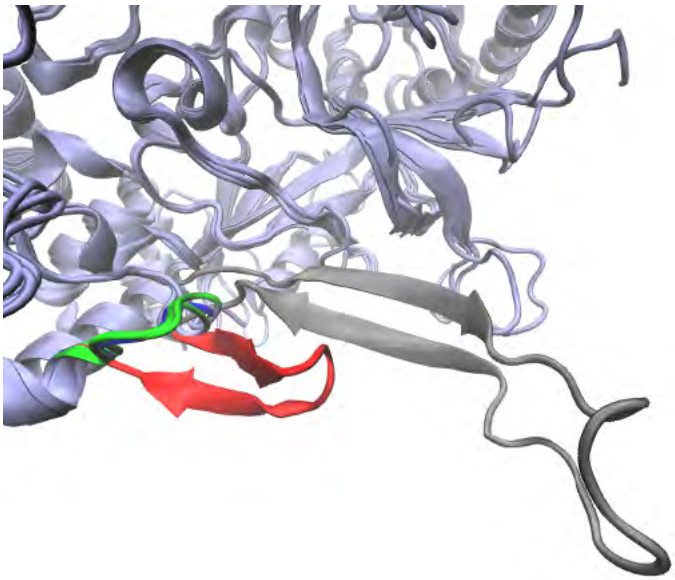
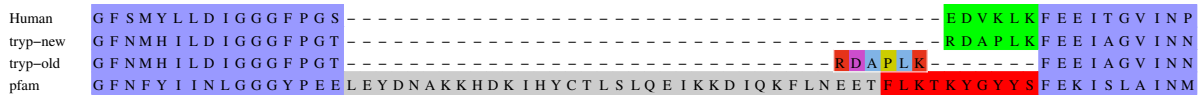


Figure 3.16: Alignment for modeling ODC. The column with the red border indicates the original position of the first residue used for modeling. The group in the green border was initially aligned eight residues upstream (Birkholtz *et al.*, 2003), and subsequently re-adjusted to make the first parasite-specific insert contiguous (Fig. 3.17). The *N*-terminus was included since it was included in more recent template structures. Amino acid similarity is indicated using the CLUSTAL (Thompson *et al.*, 1997) colour scheme.



(a) Effect of old and new ODC modeling alignments.



(b) Old and new ODC modeling alignments.

Figure 3.17: Effect of alignment on O_1. The core structures of the templates and models according to old and new alignments are all in light-blue. In the original alignment used to model *P. falciparum* ODC (Birkholtz *et al.*, 2003) an insertion (red) is introduced into a well-conserved part of both templates (green). In the models build according to the new alignment, no insertion is introduced (blue). The complete structure for O_1 from the old model is coloured red and grey. O_1 was omitted in the new model. Furthermore, generating models using the old alignment also yielded structures with knots, and on examining the likely position of this sub-insert, it appeared that it would be better placed if made contiguous.

3.3.1.2 Validation of models

PROCHECK (MORRIS *et al.*, 1992) and WHATIF (version 20051209-2240, Vriend 1990) were both used to determine the quality of the models. While it is not possible to correct all modeling errors, "better" models can be selected by use of global scores and knowledge pertinent to the protein. Both PROCHECK and WHATIF provide global scores that can be used to discard poorer models. From WHATIF, the RMS Z-scores were plotted together with G-factor overall average from PROCHECK. For both scores a larger value indicates a better model. RMS Z-scores greater than zero are better than average, while the overall average of the G-factor should be larger than -0.5. A second plot was used for the structure Z-scores from WHATIF. These values should be close to one. These plots were then used to select five models for protein-protein docking. It was observed that the cluster models tended to score highly for many of the protein quality checks and were thus added to the list of five. Thus, for each *Plasmodium* species six models each of AdoMetDC and ODC were used for protein-protein docking.

The modeling process only generated heavy atom models of the proteins involved. Protein-protein docking does not usually make use of full-atom models with all hydrogens included. Thus the models were used as is without addition of hydrogens or further refinement such as minimisation.

3.3.1.3 Docking of AdoMetDC/ODC with 3D-DOCK/FTDOCK

FTDOCK2 (Jackson *et al.*, 1998) was used to dock each AdoMetDC model against each ODC model within each *Plasmodium* species. Thus 36 docking runs were performed for species, resulting in 180 runs in total. In each docking run, AdoMetDC was treated as mobile while the ODC dimer was kept static. An angle step of 10° was used.

As a control the docking was repeated for all current human structures. The AdoMetDC PDB ids are: 1I72, 1I79, 1I7B, 1I7C, 1I7M, 1JEN, 1JLO and 1MSV. In the case of dimeric structures (1I7M, 1JEN, 1JLO and 1MSV), the dimers were split and each monomer was used for docking. The ODC structures used were 1D7K and 2OO0, both dimeric. Using all the possible AdoMetDC/ODC combinations gave 24 runs.

The docking results were re-scored using RPSCORE from FTDOCK and the top 100 hits from each docking run analysed. The top 100 results from each run were combined and superimposed based on the ODC co-ordinates using VMD.

The xeon cluster of the BMIC (Bio-Medical Informatics Centre) at the Meraka Institute, CSIR was used to perform the dockings in parallel.

3.3.1.4 Analysis

FTDOCK scores orientations based on complementarity and electrostatic interactions. These results were then re-scored using RPSCORE of 3D-DOCK which uses a pairwise residue scoring system derived empirically from a database of known protein-protein interfaces. For each *Plasmodium* species the C_α co-ordinates of the AdoMetDC-ODC dimer complex of the top 100 orientations from each docking run were extracted. All 3600 orientations were superimposed based on the ODC co-ordinates. The superimposition was performed either by minimising RMSD of the $C_\alpha$positions or using maximum likelihood superimposition as implemented by the THESEUS (Theobald and Wuttke, 2006b) program. The later method accounts for correlation and unequal variances (Theobald and Wuttke, 2006a).

3.4 Results and discussion

3.4.1 modeling of AdoMetDC/ODC

3.4.1.1 Model quality

From the 100 models generated for each domain in each species a smaller set was selected for protein-protein docking experiments. Model generation inherently produces ”errors” where certain structural features deviate substantially from the norm. For the current purpose it

was impractical to correct all of these, therefore a set of models was chosen using PROCHECK and WHATIF that minimises the number of errors. For selecting models the RMS Z-score, Structure Z-score (WHATIF) and overall G-factor (PROCHECK) were used. The RMS Z-score includes bond lengths, bond angles, ω-angle restraints, side chain planarity, improper dihedral distribution, inside/outside distribution and should be close to 1. The structure Z-score includes 1st generation packing quality, 2nd generation packing quality, Ramachandran plot appearance, χ_1/χ_2-rotamer normality and backbone conformation, where positive is better than average. The PROCHECK G-factor was included with the RMS Z-scores because it is in the same order of magnitude. The G-factor should be > -0.5 for a good quality structure. While this is typical for experimental structures, the score is usually in this range for homology models. The factors measured by the G-factor overlap with the WHATIF Z-scores. This is apparent in that a higher G-factor usually correlates with good Z-scores. The results for *P. berghei* are included as a representative example (Fig. 3.18, Fig. 3.19). For completeness the scores for the other species are included in Appendix B.1. From the series of 100 models, five were chosen based on their scores. The cluster[2] models were generally observed to score highly, and were therefore included to bring the total up to six models for each domain included for docking. The list of models chosen from the 100 generated for each domain is given in Table 3.1.

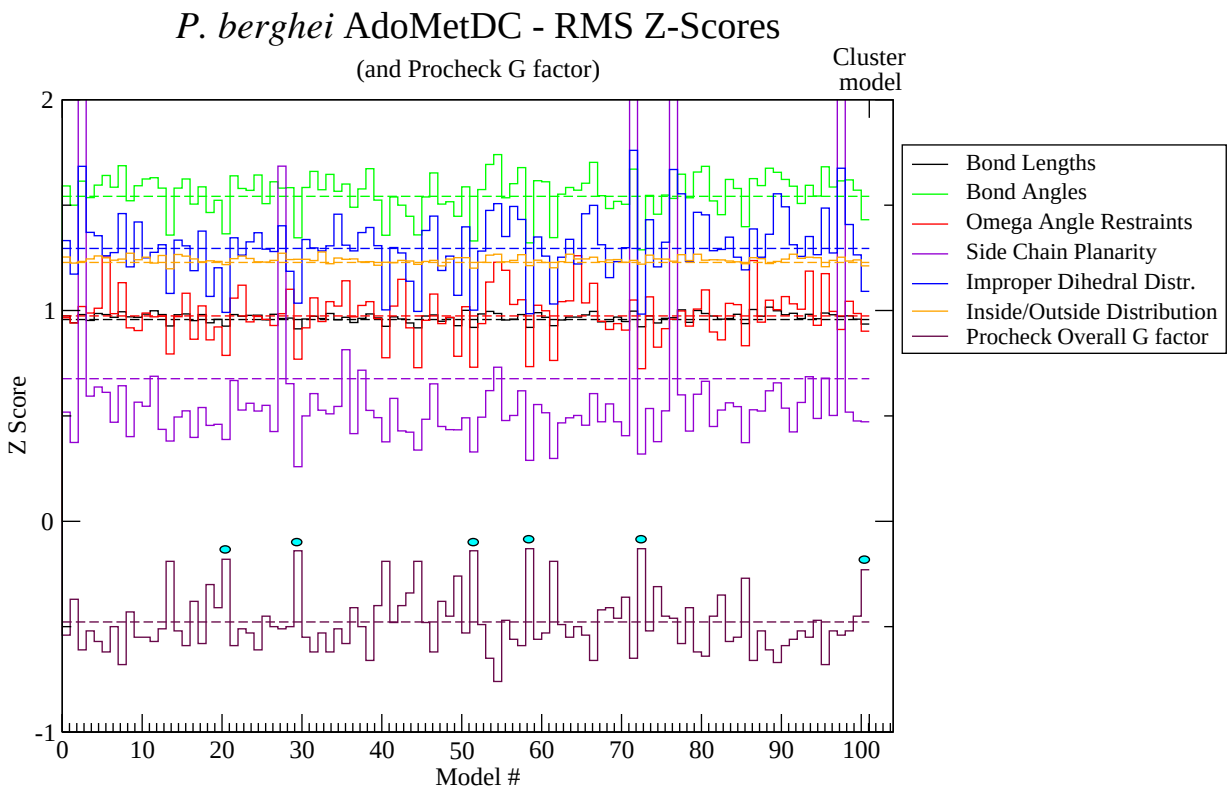


Figure 3.18: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. berghei* AdoMetDC. The models chosen for docking are indicated by the cyan dots. Averaged values are indicated by the dashed lines.

[2]The cluster model is an averaged structure derived from the 100 models built.

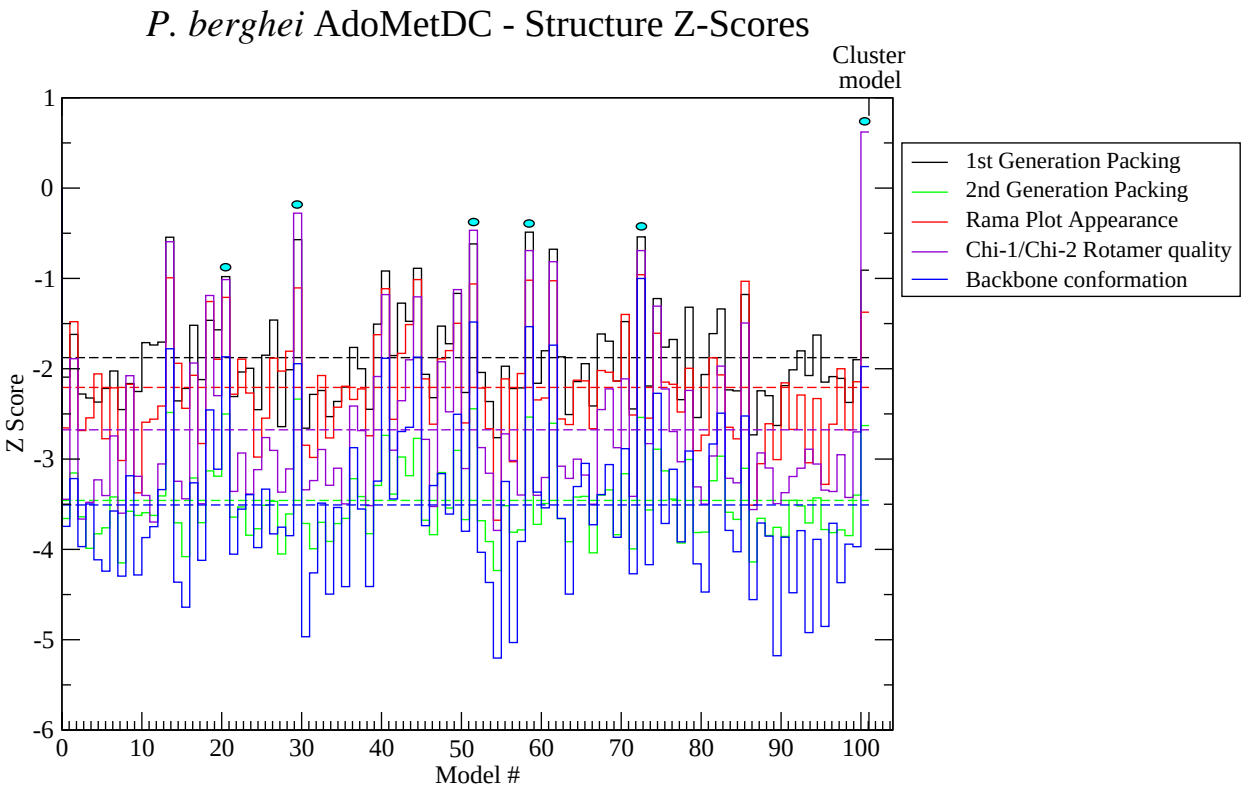


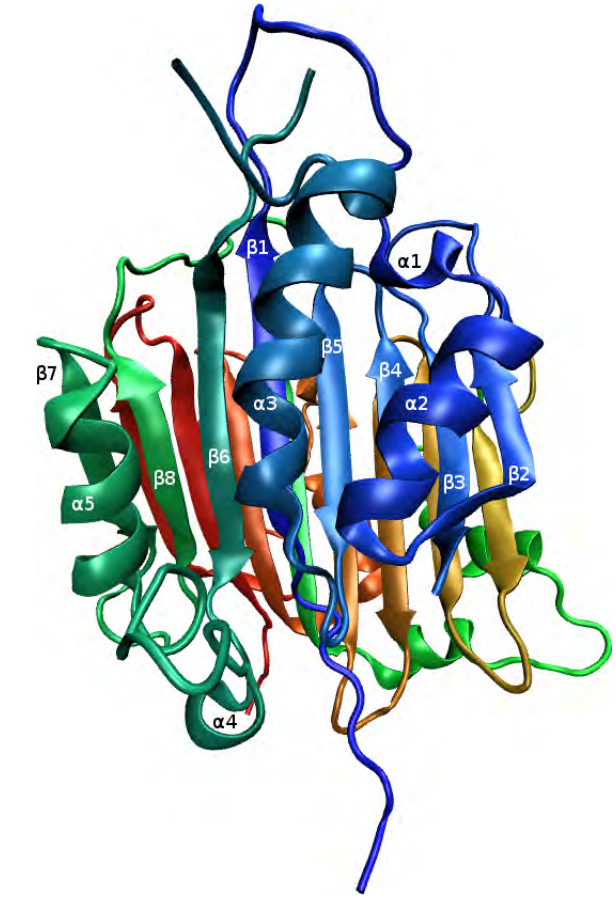
Figure 3.19: WHATIF Structure Z-scores for modeling of *P. berghei* AdoMetDC. The models chosen for docking are indicated by the cyan dots. Averaged values are indicated by the dashed lines.

Table 3.1: Model numbers chosen for protein-protein docking. The associated WHATIF and PROCHECK G-factors are given in Fig. 3.18, Fig. 3.19 and App B.1.

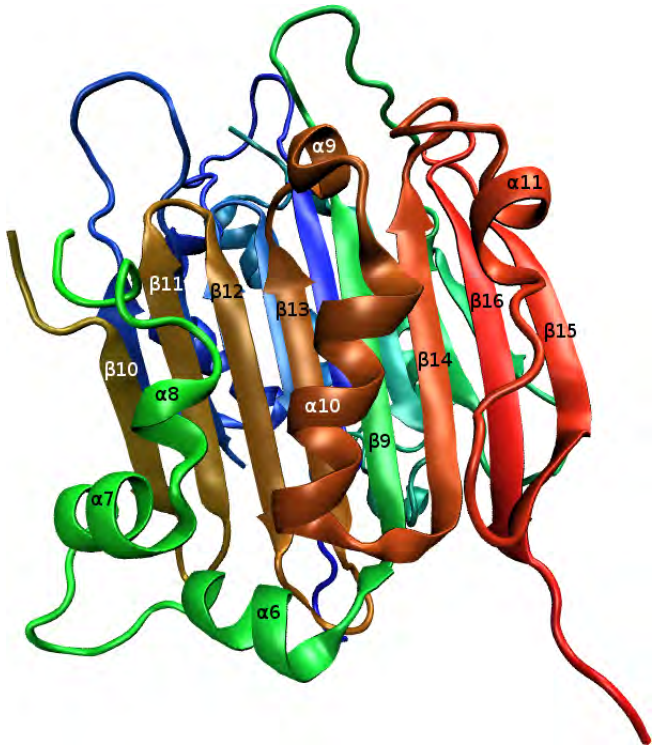
Species	AdoMetDC	ODC
P. berghei	21, 30, 52, 59, 73	6, 13, 23, 32, 58
P. falciparum	17, 32, 56, 57, 95	12, 18, 23, 44, 75
P. knowlesi	1, 3, 13, 32, 39	26, 32, 39, 60, 76
P. vivax	23, 39, 54, 69, 84	3, 43, 54, 56, 75
P. yoelii	4, 47, 59, 63, 69	6, 58, 76, 78, 98

3.4.1.2 Topology and tertiary structure of AdoMetDC/ODC models

In the results that follow reference is often made to the topology and tertiary structures of AdoMetDC and ODC. The typical AdoMetDC and ODC model topologies and tertiary structures are therefore depicted in Figures 3.20 and 3.21 for future reference.

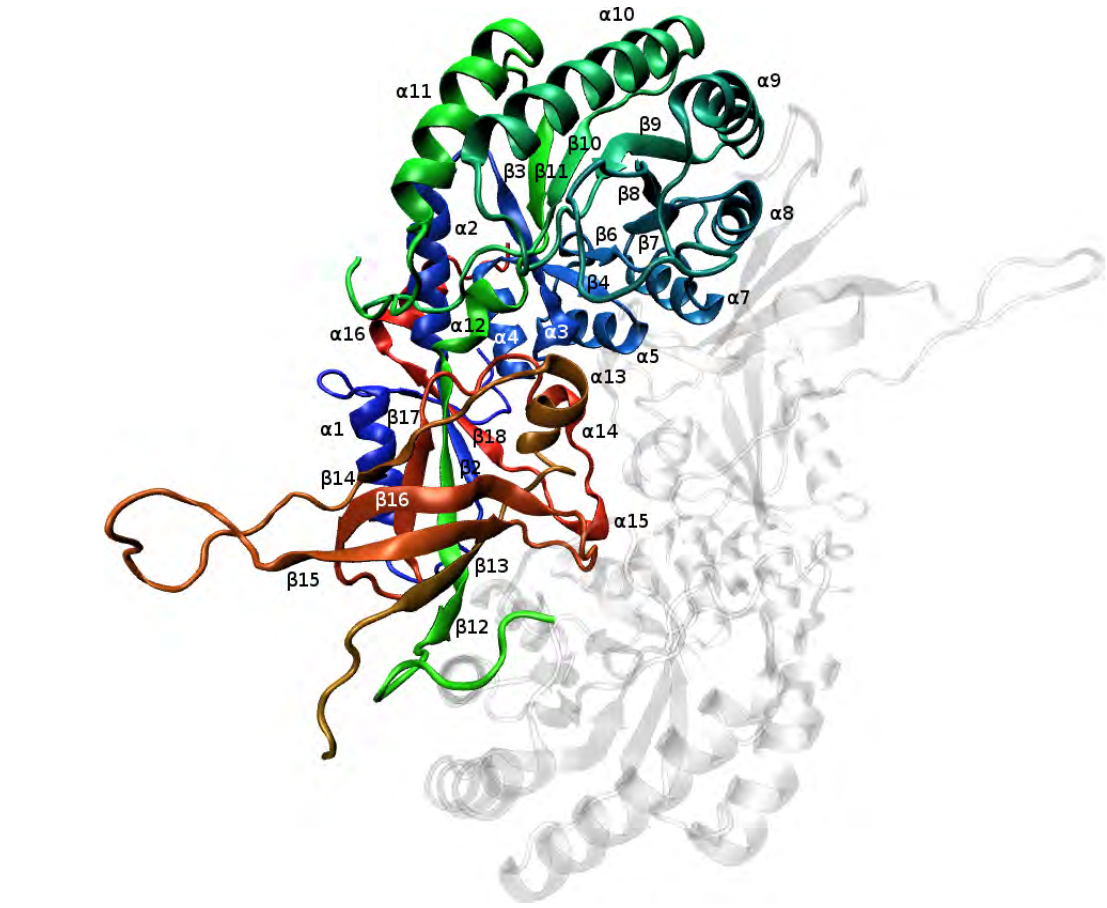


(a) Top

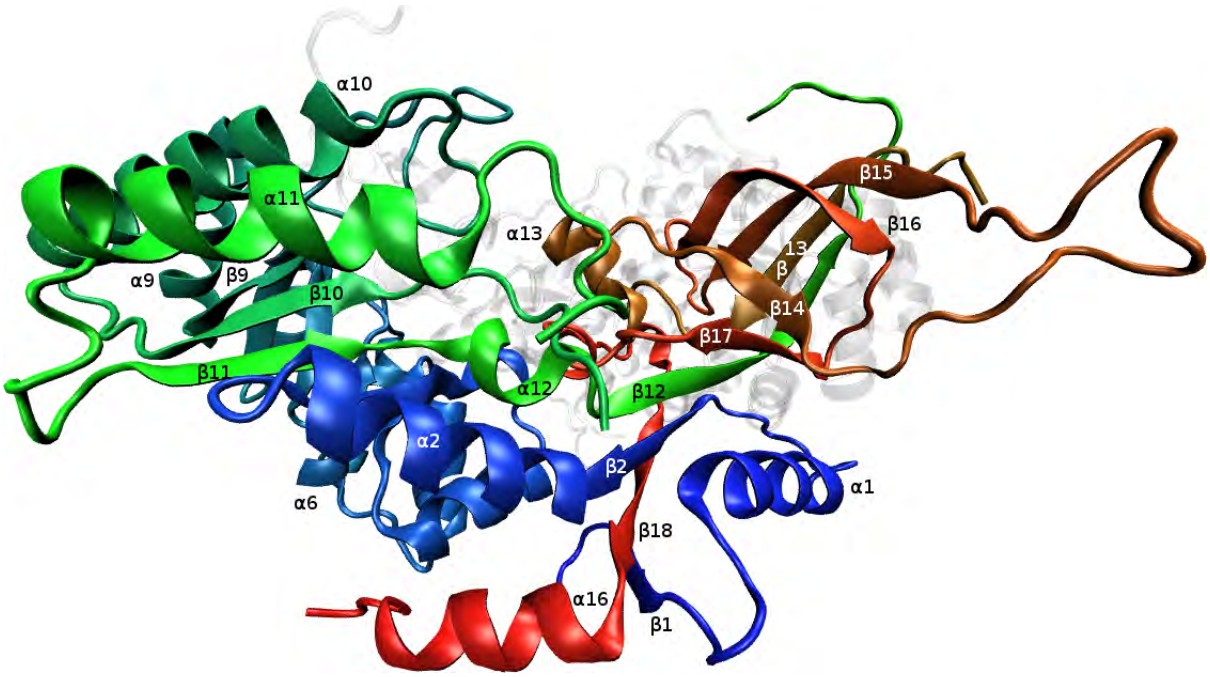


(b) Bottom

Figure 3.20: Tertiary structure of the *Plasmodium* AdoMetDC model. The cluster model for *P. falciparum* is depicted. Secondary structure elements (βx for strands, αx for helices) are indicated. Increasing sequence position is indicated by a blue→green→red colour gradient.



(a) Top



(b) Side

Figure 3.21: Tertiary structure of the *Plasmodium* ODC model. The cluster model for *P. falciparum* is depicted. Secondary structure elements (βx for strands, αx for helices) are indicated. Increasing sequence position is indicated by a blue→green→red colour gradient. One monomer is "ghosted".

3.4.1.3 Definition of poses

For the purpose of analysing docking results it is useful to define poses of both AdoMetDC and ODC that cover the entire surface of the respective protein or protein complex. For each domain six roughly orthogonal poses are arbitrarily defined (Fig. 3.22 and 3.23). While AdoMetDC is asymmetric, within the ODC dimer sides 1 and 2 and sides 3 and 4 are related by two-fold symmetry.

(a) Top

(b) Bottom

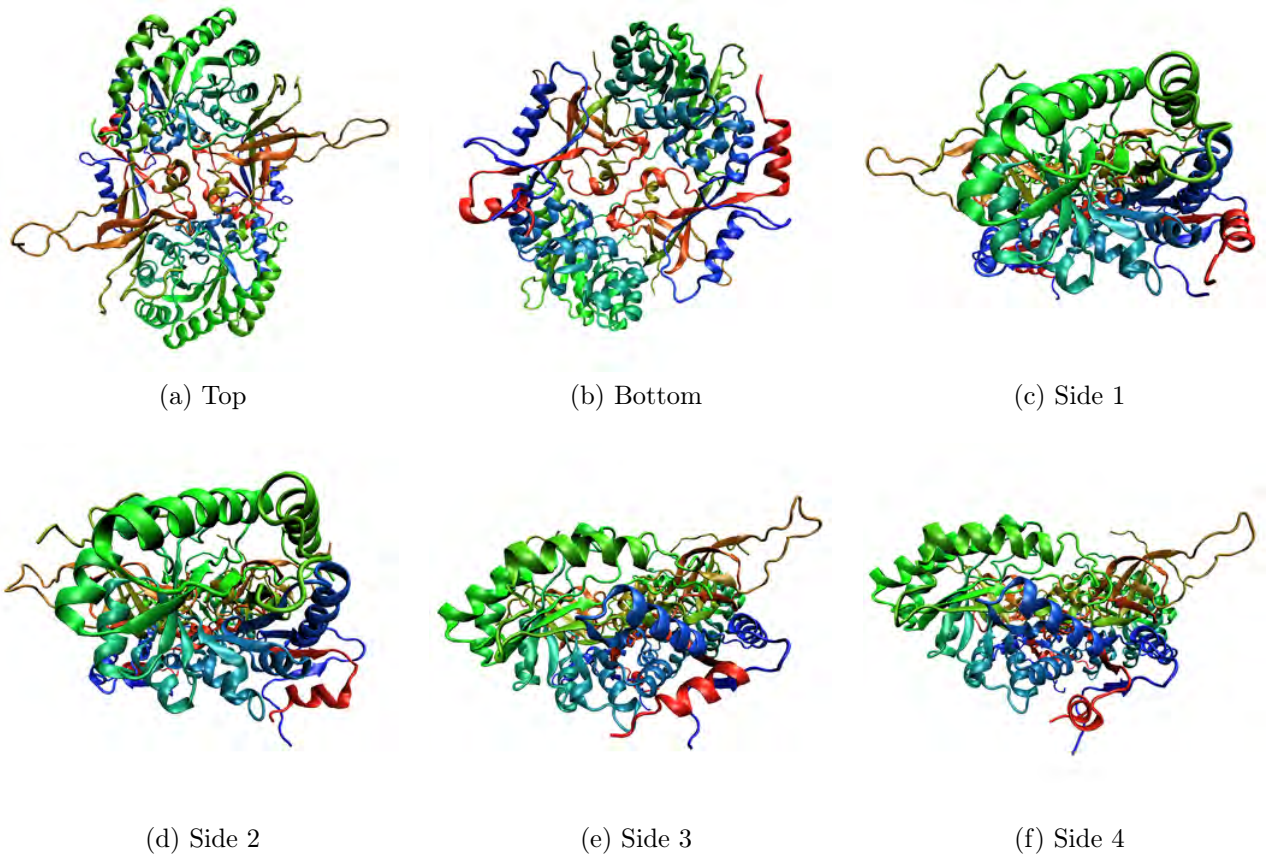
(c) Side 1

(d) Side 2

(e) Side 3

(f) Side 4

Figure 3.22: Arbitrarily defined orthogonal poses for AdoMetDC. Increasing sequence position is indicated by a blue→green→red colour gradient.



(a) Top (b) Bottom (c) Side 1

(d) Side 2 (e) Side 3 (f) Side 4

Figure 3.23: Arbitrarily defined orthogonal poses for ODC. Increasing sequence position is indicated by a blue→green→red colour gradient.

3.4.1.4 Surface distribution of divergence

The conservation of residues on the surface of the models was analysed for several reasons. Firstly, since bifunctional AdoMetDC/ODC only exists in *Plasmodium sp.* it could be expected that there would be greater divergence on the contacting surfaces of each domain compared to non-*Plasmodium* cognates. Also, if the various *Plasmodium* models are compared with each other, it is expected that there should be greater conservation between the contacting surfaces (Yan *et al.*, 2008).

When comparing the six poses (Defined in Fig. 3.22) of AdoMetDC certain faces of the protein are more conserved according to the alignment used for modeling (Fig. 3.24-3.25 and Appendix B.2). In particular, the top face is highly conserved across all the species modelled, with greater conservation between more closely related species. This is observed between *P. berghei*, *P. yoelii* which both infect rodents, and to a lesser degree between the primate infection species of *P. falciparum*, *P. knowlesi* and *P. vivax* (in particular the latter two). The top face also corresponds to the site of proteolytic cleavage and is thus closer to the active site. In contrast, the bottom face shows a high degree of divergence, except between the very closely related *P. berghei* and *P. yoelii*. There is a reasonably high degree of conservation on side 2 of AdoMetDC, followed by side 3. These trends are observed when comparing species pairwise or against all remaining *Plasmodium* species. The AdoMetDC

inserts A_1, A_2 and A_3 are predicted to extend from side 1, the top face and the bottom face respectively. Thus the most divergent insert, A_3 is also predicted to extend from the least conserved face of AdoMetDC. The results for *P. berghei* are included as a representative example (Fig. 3.24, Fig. 3.25). The results for the other species are included in Appendix B.2.

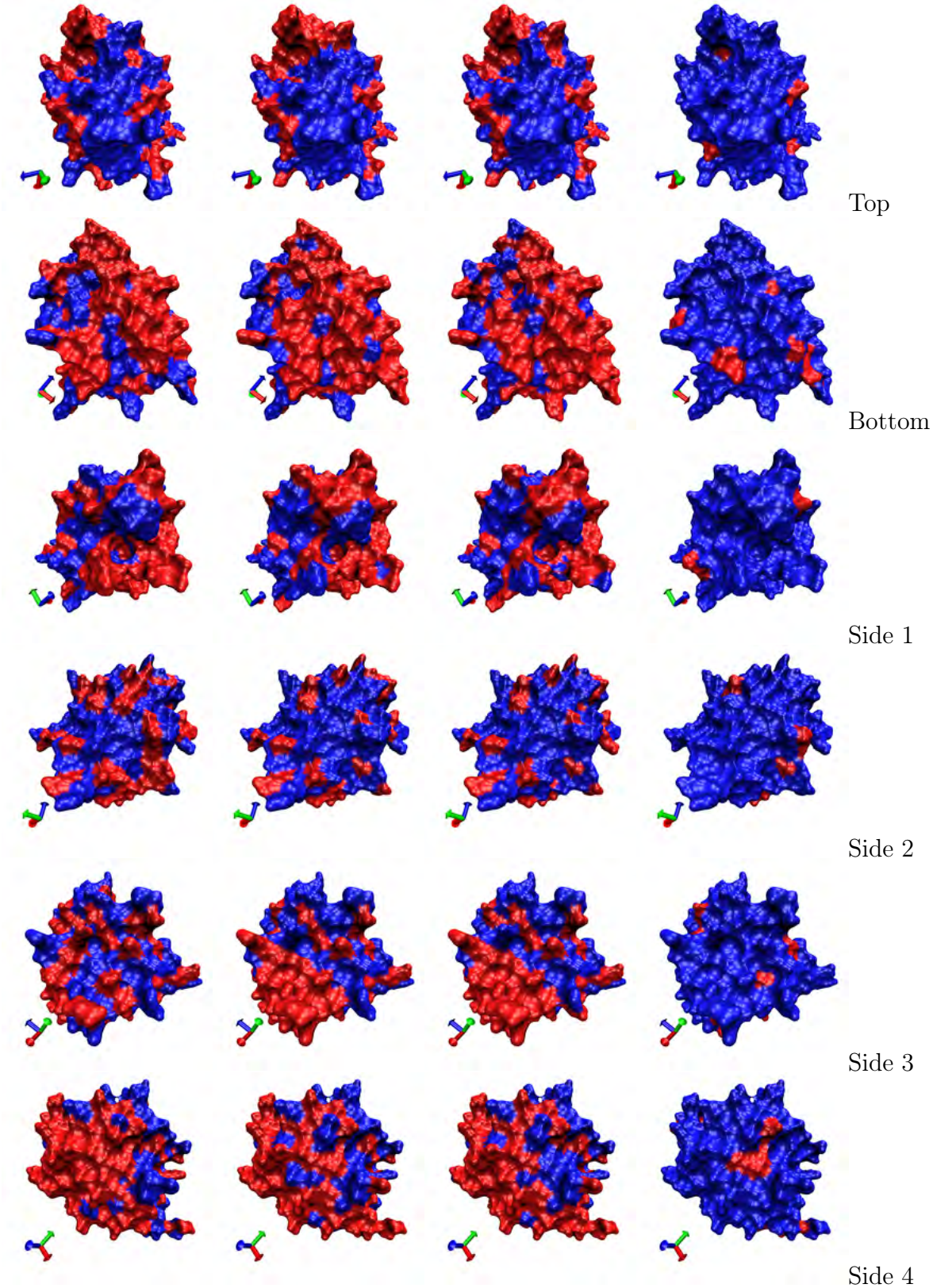


Figure 3.24: Pairwise conservation of *P. berghei* AdoMetDC surface residues. *P. berghei* is compared to *P. falciparum*, *P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1-6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

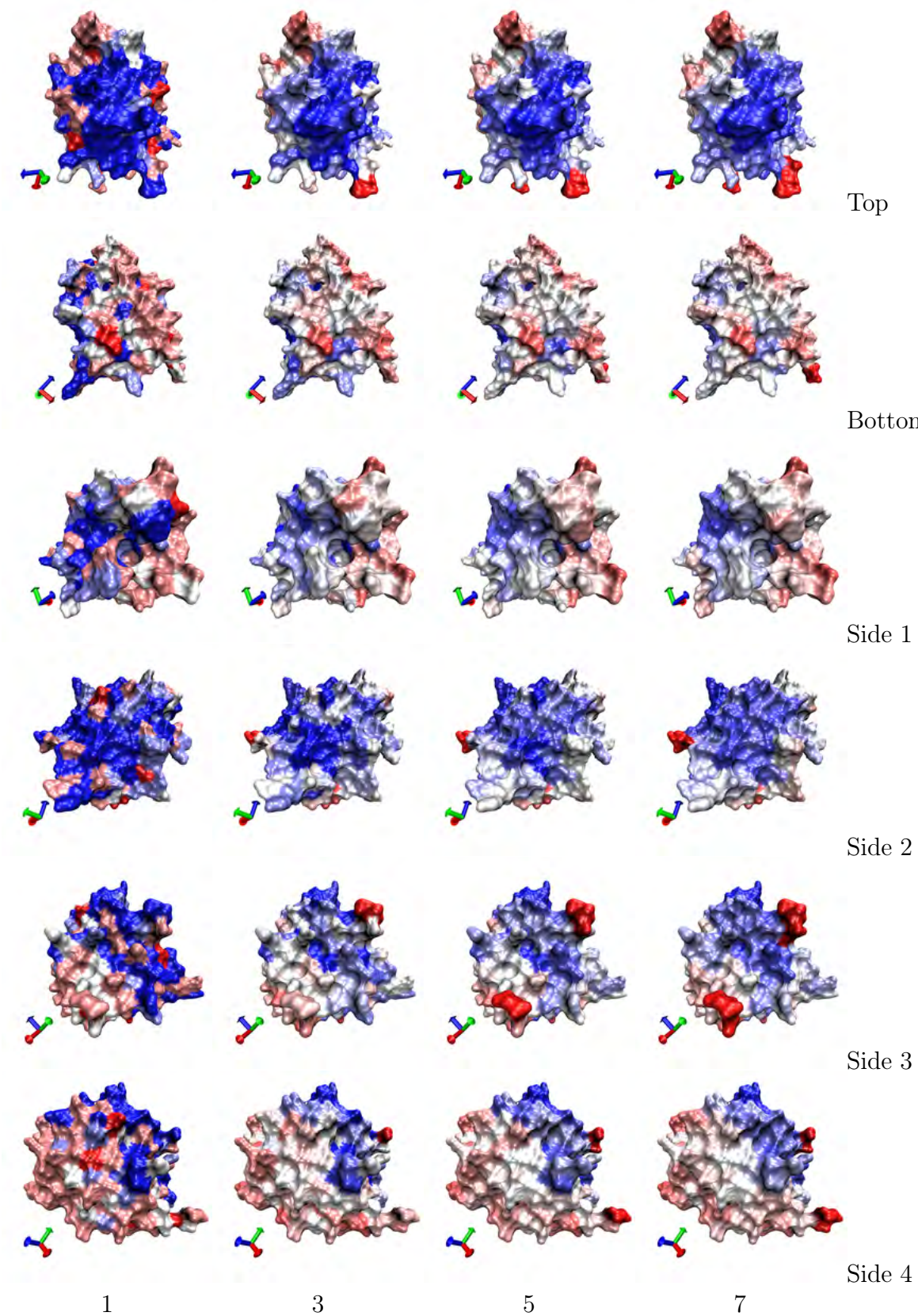


Figure 3.25: Conservation of *P. berghei* AdoMetDC surface residues. *P. berghei* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

Recombinantly expressed *P. falciparum* AdoMetDC is known to dimerise, as does human AdoMetDC. By homology with human AdoMetDC the site of association should be side 4, however according to the models there is a high degree of divergence. Therefore, because interfaces are expected to be more conserved, *Plasmodium* AdoMetDC may dimerise via a different interface.

Comparing the six poses of ODC (Defined in Fig. 3.23) reveals that the top face is highly conserved across all *Plasmodium* species compared to the bottom face (Fig. 3.26,Fig. 3.27). The top face also corresponds to where the active sites are located. It is also worth noting that both ODC inserts are also predicted to extend from the top face. None of the other faces stands out is being particularly well conserved or highly divergent. Sides 1 and 2 display essentially the same pattern, as do sides 3 and 4 due to two-fold symmetry of the ODC dimer (i.e. identical primary structure). The slight differences observed are due to the small conformational differences of the protein surface. For all faces there is higher conservation between *P. berghei* and *P. yoelii* and between *P. knowlesi* and *P. vivax*. These trends are observed when comparing species pairwise or against all remaining *Plasmodium* species. The results for *P. berghei* are included as a representative example. The results for the other species are included in Appendix B.2.

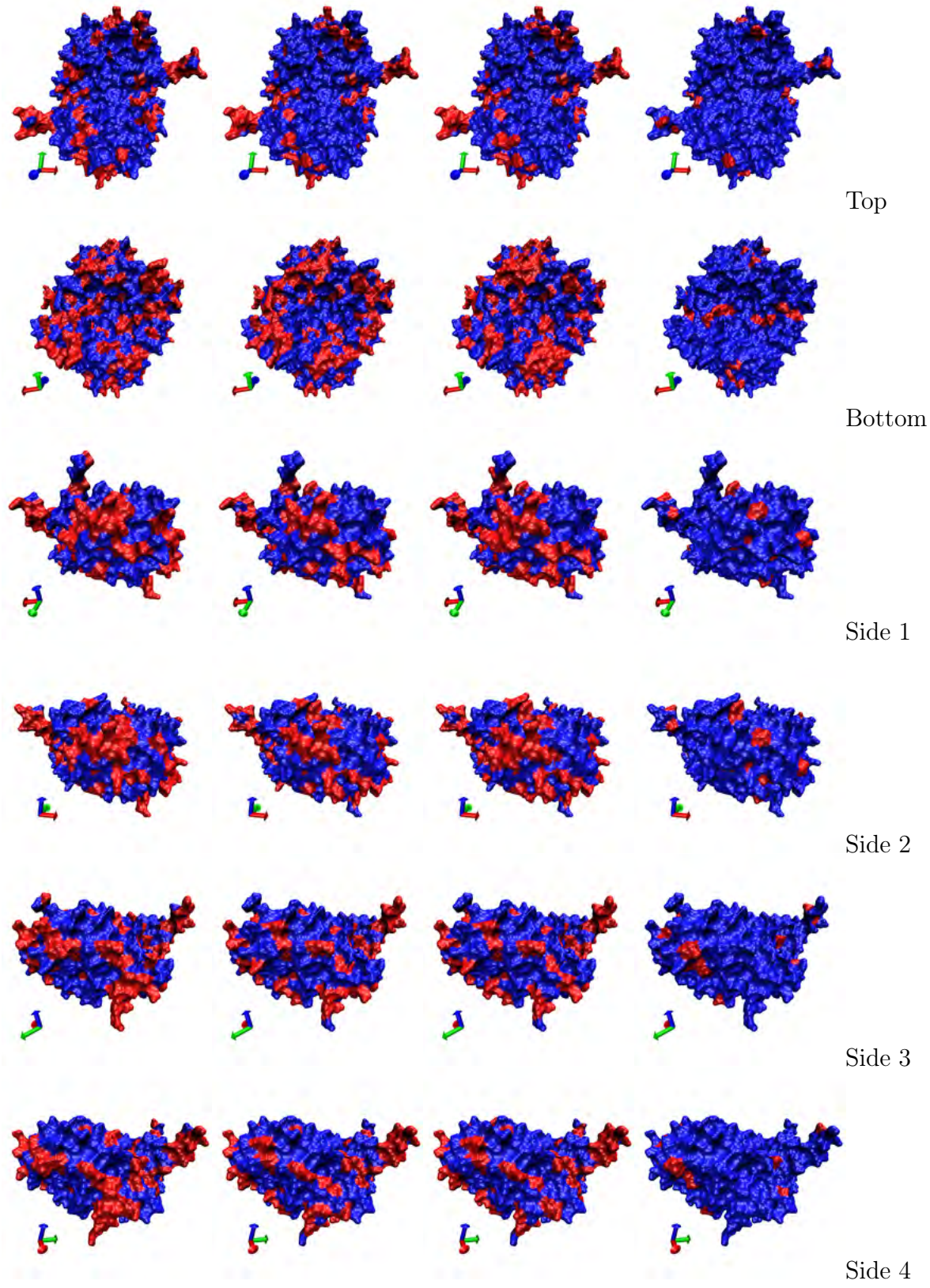


Figure 3.26: Pairwise conservation of *P. berghei* ODC surface residues. *P. berghei* is compared to *P. falciparum*, *P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

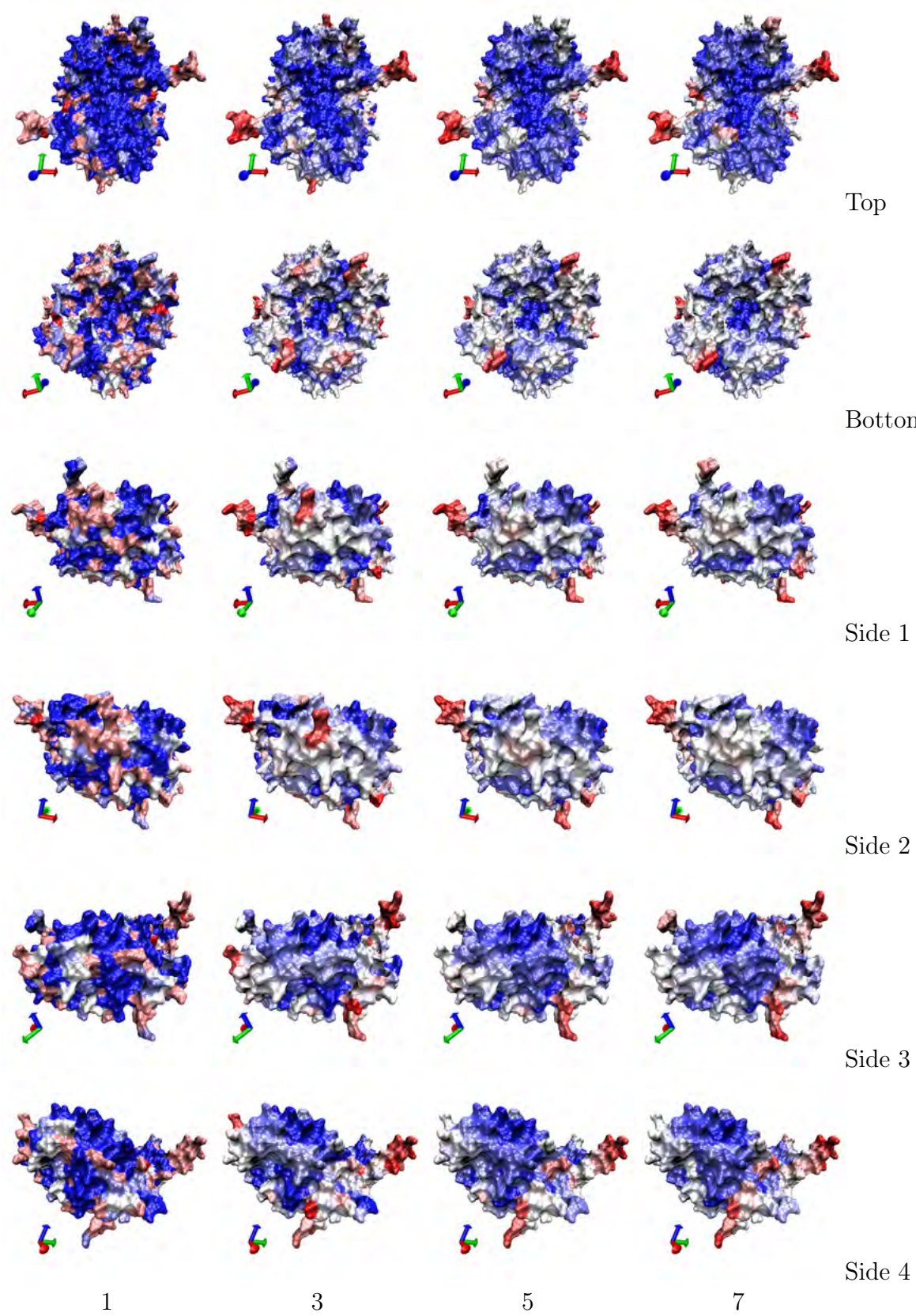


Figure 3.27: Conservation of *P. berghei* ODC surface residues. *P. berghei* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

3.4.2 Docking of AdoMetDC/ODC

3.4.2.1 Docking scores

The scores for the docking of *Plasmodium* AdoMetDC/ODC vs the human cognates appear substantially higher. To determine whether this was statistically significant, Wilcox rank sum tests were performed on the top 100 RP (Residue pairing potential) scores as well as the top 10 000. The Wilcox rank sum test was used to determine whether docking scores between different AdoMetDC and ODC pairs within the same species are from the same distribution. These tests were carried out using the R program (http://www.r-project.org/) for scores within a species (different docking runs compared against each other) as well as between the human and various *Plasmodium* species.

According to this test between 14 and 33% of possible pairs belong to the same distribution (5% level significance) including human structures and *Plasmodium* models (Table 3.2). This was both for the top 100 and all 10 000 dockings. According to this test the scores from most pairs of runs are from different distributions. The Wilcox rank sum scores comparing *Plasmodium* and human runs are extremely small (in the order of 10^{-34}) and therefore extremely unlikely to be from the same distribution. When comparing runs within a species, some pairs are predicted to be from the same distribution (≥ 0.05), but not all. Thus, although the scores for the *Plasmodium* models are significantly different from the human structures, this is also true for different docking runs within a species. Thus the Wilcox rank sum test is insufficient to declare the difference between *Plasmodium* and human significantly different.

Table 3.2: Significant Wilcox rank sum tests. RP scores from docking runs within the same species were compared against each other (pairwise). The scores were considered to come from the same distribution if the p-value for the Wilcox rank sum test was ≥ 0.05.

Species	Same distribution (s)	Total pairings (t)	Ratio (s/t)
Human (Top 100)	61	276	0.22
Human (All 10 000)	39	276	0.14
P. berghei (Top 100)	160	630	0.25
P. berghei (All 10 000)	162	630	0.26
P. falciparum (Top 100)	198	630	0.31
P. falciparum (All 10 000)	206	630	0.33
P. knowlesi (Top 100)	210	630	0.33
P. knowlesi (All 10 000)	156	630	0.25
P. vivax (Top 100)	122	630	0.19
P. vivax (All 10 000)	144	630	0.23
P. yoelii (Top 100)	170	630	0.27
P. yoelii (All 10 000)	171	630	0.27

The scores of the various docking runs were also analysed by means of box and whisker plots in R (R Development Core Team, 2009). So-called "notches" (Fig. 3.28) were included

which can be used to determine whether the medians from two samples differ. If the notches between two samples do not overlap, this is strong evidence that the medians differ, i.e. the samples are drawn from different underlying distributions. The overlaps of all notches were then compared within the human and *Plasmodium* scores. A single overlap was observed for all *Plasmodium* species, while two overlaps were observed for human. The two human overlaps were separated by a very small difference and it is expected will have formed a complete overlap when run with a set of models derived from different random number seeds. A large difference is observed between the human and *Plasmodium* overlaps for all five *Plasmodium* species (Fig. 3.28 and Appendix B.3). It therefore appears unlikely that a single overlap would be observed between human and *Plasmodium* using a different set of models. From these results the top 100 RP scores between *Plasmodium* and human were judged to be significantly different. Furthermore, because the scores from *Plasmodium* were significantly higher than the human scores this suggests that *Plasmodium* AdoMetDC/ODC are more likely to form a physical complex than human AdoMetDC/ODC. This is obviously in agreement with experiment, since *Plasmodium* AdoMetDC/ODC occur in a bifunctional protein, while human AdoMetDC and ODC have never been reported to physically interact (Birkholtz *et al.*, 2004).

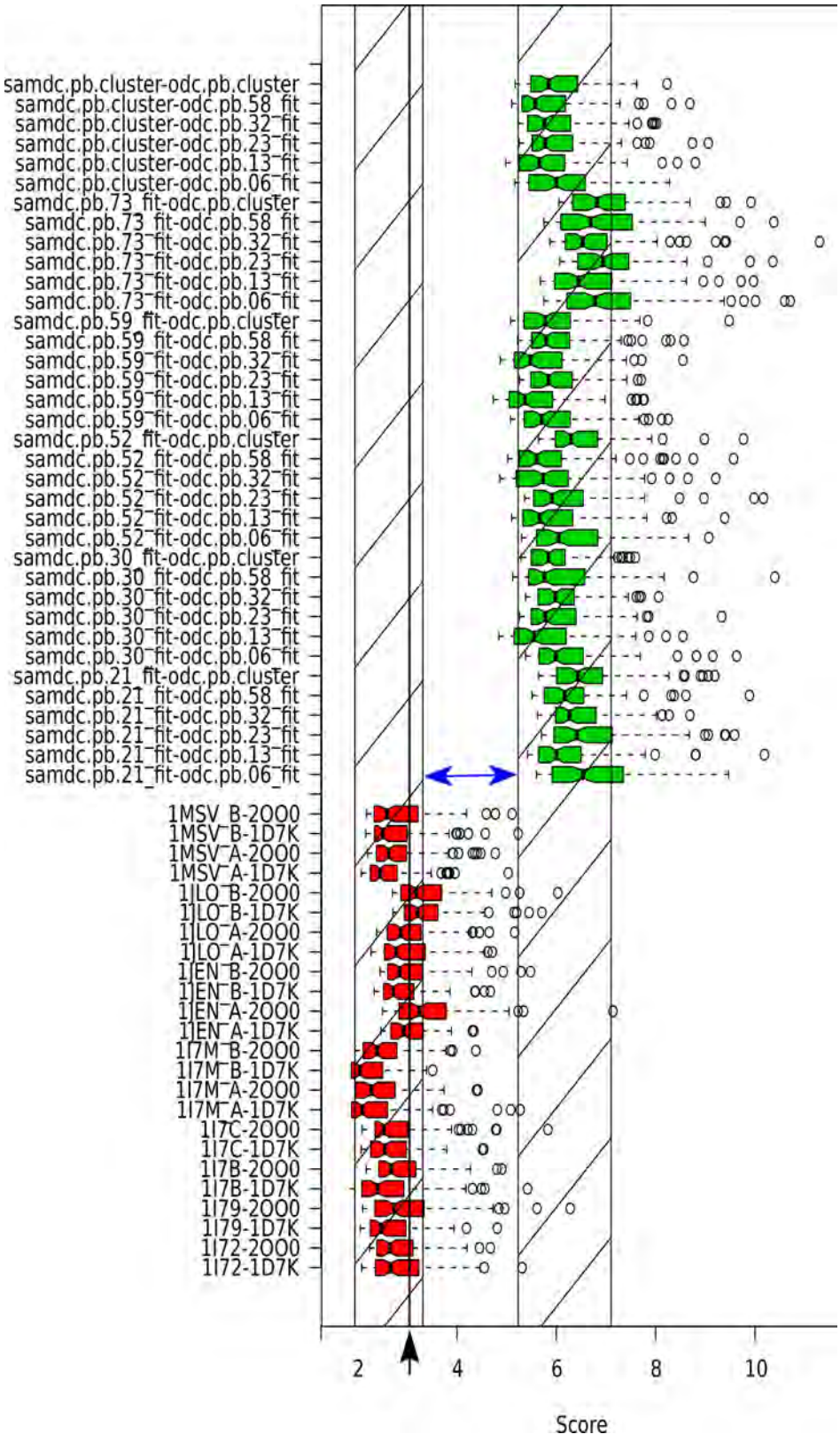


Figure 3.28: Distribution of top 100 RP scores for docking of human (red) and *P. berghei* (green) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human with a gap separating them (black arrow). The gap is too small to be visible on this scale. One overlap is visible for *P. berghei*. A substantial gap exists between the human and *P. berghei* overlaps (blue arrow), suggesting the distribution of scores is significantly different between human and *P. berghei*. The FTDOCK RP score is plotted on the x-axis against the AdoMetDC/ODC model (green) or structure (red) combinations. Similar considerations apply to the other *Plasmodium* species (See Appendix B.3).

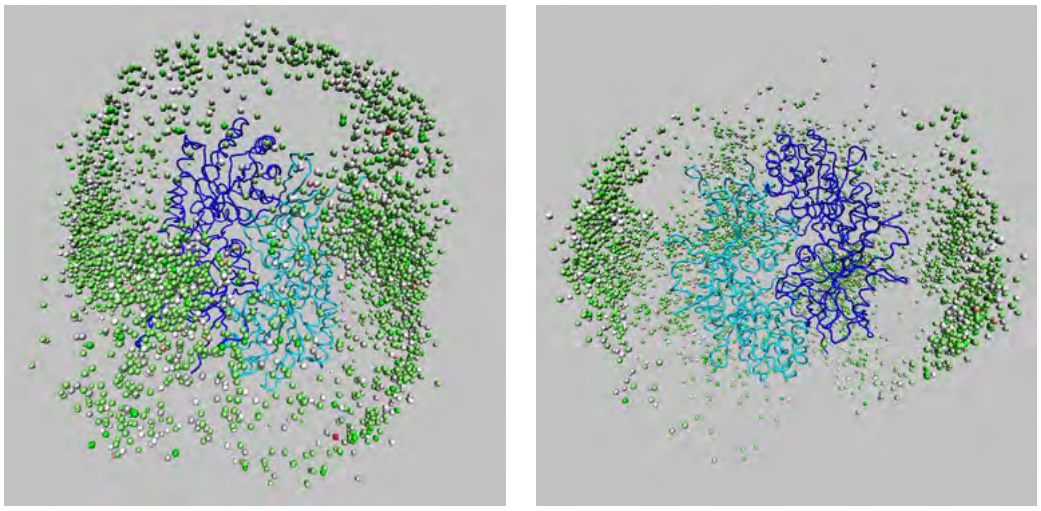
3.4.2.2 Centre of mass distribution

The distribution of the centre of mass (COM) of each domain was analysed relative to the partner protein. An uneven distribution was obtained for both AdoMetDC and ODC for the top 100 predictions across all five *Plasmodium* species. Certain trends were also observed when comparing the various species. In all species AdoMetDC was predicted to favour the top face of ODC compared to the bottom face (Fig. 3.29, Fig. 3.30, Section B.4[3]). Furthermore, the top face interactions are more concentrated around the β-sheet domain compared to the TIM-barrel domain. This is particularly noticeable when viewing the distribution from the bottom face. A large distribution of COMs is also observed down the sides of the ODC dimer, also favouring the β-sheet domain, making contact with α1, α16, β2 and α2. This is especially noticeable in *P. berghei*, *P. knowlesi*, *P. vivax* and *P. yoelii*. In *P. falciparum*, however far fewer COMs are found in this region. Instead, the COMs for *P. falciparum* are concentrated far more on the top face compared to the other *Plasmodium* species with far fewer interactions in the side faces compared to the other *Plasmodium* species. Compared to the distribution of divergence there is also broad agreement in that there is both a high density of COMs and greater sequence conservation on the top face. Sequence divergence for the predicted interacting regions on sides 1 to 4 is only moderate, however.

The prediction of AdoMetDC-ODC interactions for sides 1 to 4 for four out of the five species is noteworthy. This would imply that the AdoMetDC domains would be too far apart to physically interact. However, recombinantly expressed AdoMetDC is observed to dimerise. Whereas if AdoMetDC binds to ODC on the top face with the sheet domain, this would also allow the two AdoMetDC monomers to make contact within the bifunctional complex. Furthermore, this would make for a more compact and energetically favourable globular complex compared to an interaction with sides 1 to 4 of the ODC dimer. It is expected that the general architecture of the AdoMetDC/ODC complex will be conserved among the various *Plasmodium* species. The low density of COMs for sides 1 to 4 in *P. falciparum* and lower sequence conservation further suggests that an interaction between sides 1 to 4 of the ODC dimer and AdoMetDC can be ruled out.

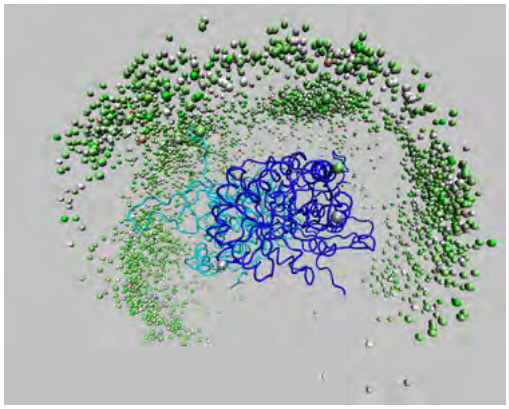
The top face of ODC also contains the active site as well as the predicted positions for the parasite-specific inserts O_1 and O_2. This confirms previous studies which demonstrate that both O_1 and O_2 affect the activity of both AdoMetDC and ODC domains . Previous results also indicate that O_1 and O_2 are required for AdoMetDC and ODC to physically interact (Birkholtz *et al.*, 2004). This would be most easily effected if the inserts themselves make contact with AdoMetDC. These results provide further support that AdoMetDC interacts with ODC via the top face.

[3] *P. berghei* is included as representative of four out of the five species, while *P. falciparum* is slightly atypical and therefore discussed further here. The remaining results are included in the Appendix (B.4)

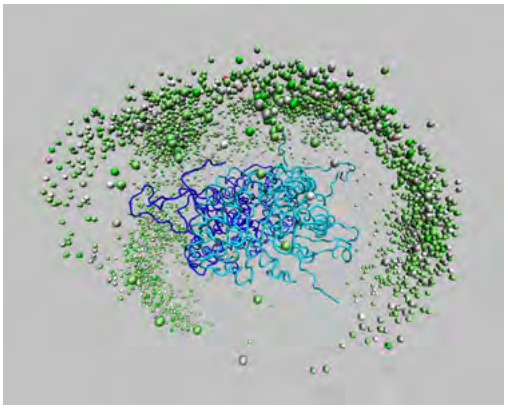


(a) Top

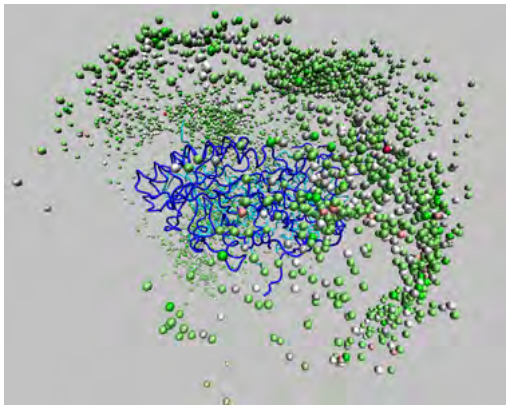
(b) Bottom



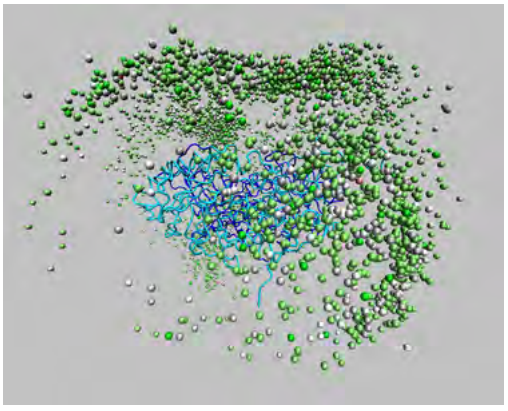
(c) Side 1



(d) Side 2

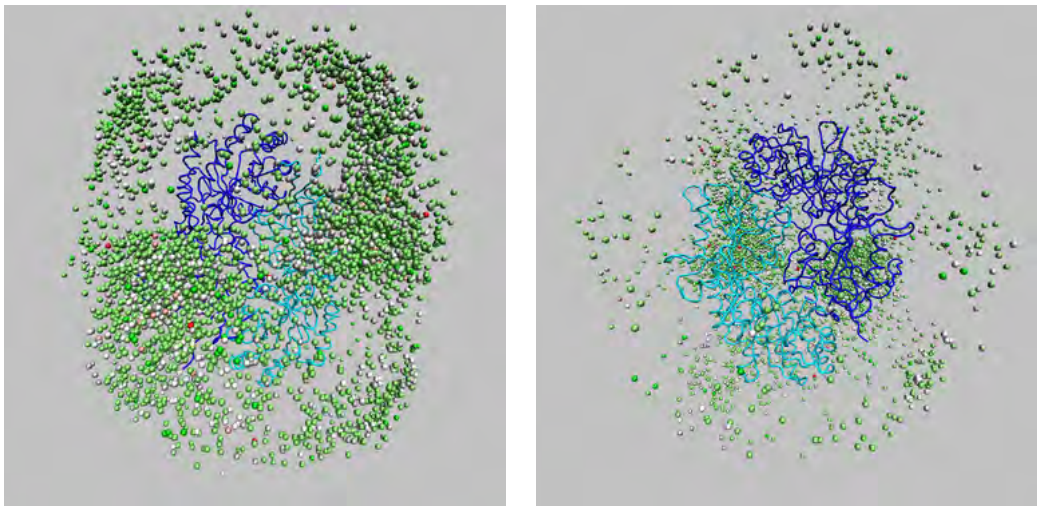


(e) Side 3



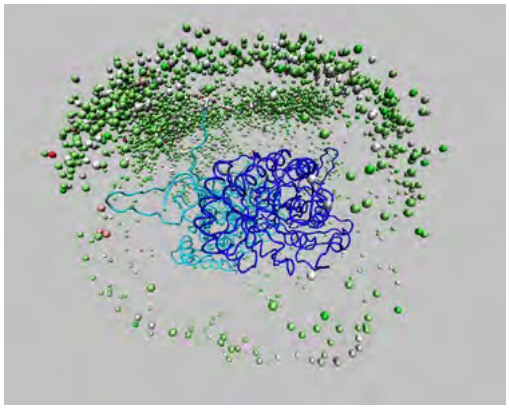
(f) Side 4

Figure 3.29: Centre of mass (COM) distributions of *P. berghei* AdoMetDC relative to ODC. ODC chains A (blue) and (cyan) B are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based on the RP score (4.73: green → 11.30: red).

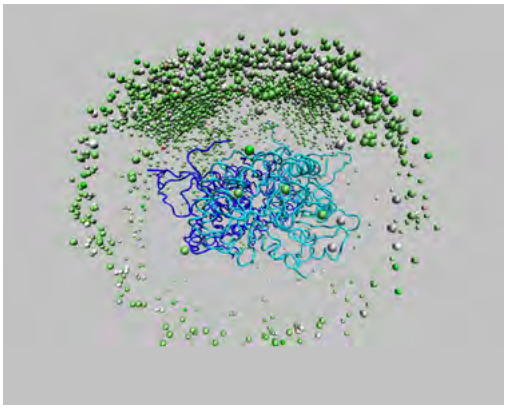


(a) Top

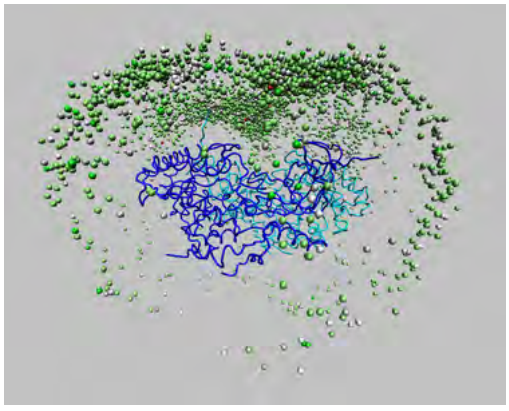
(b) Bottom



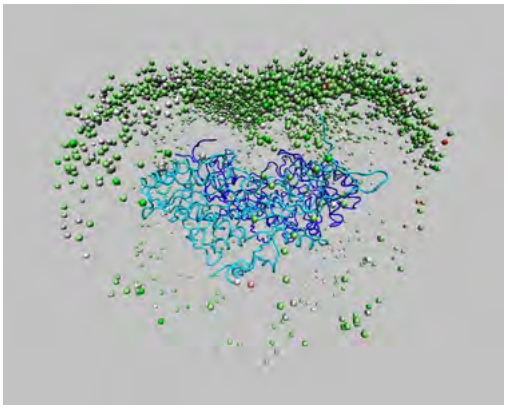
(c) Side 1



(d) Side 2



(e) Side 3



(f) Side 4

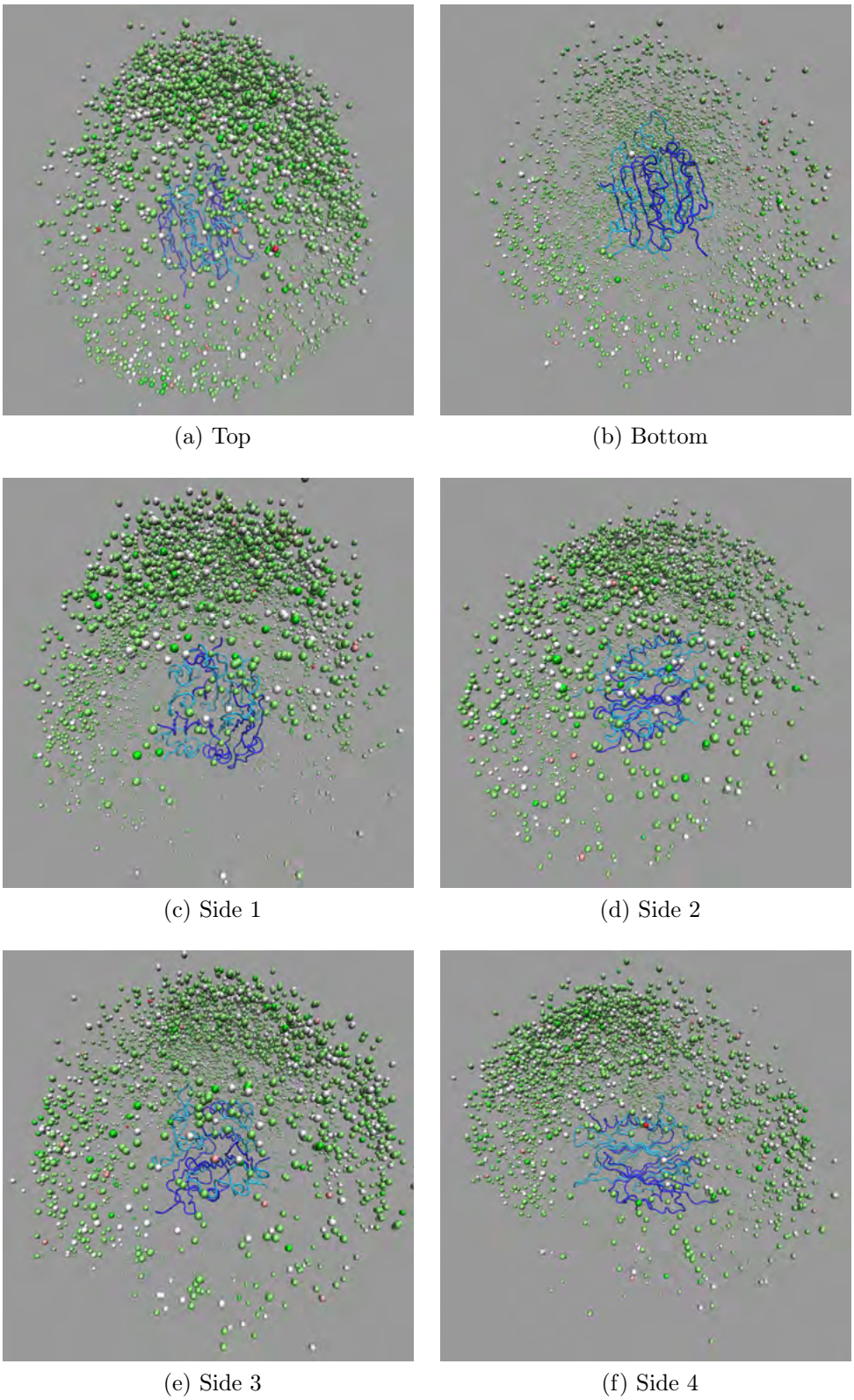
Figure 3.30: Centre of mass (COM) distributions of *P. falciparum* AdoMetDC relative to ODC. ODC chains A (blue) and (cyan) B are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based on the RP score (4.87: green → 12.64: red).

The distribution of ODC COMs relative to AdoMetDC was less consistent across the five *Plasmodium* species. The most consistent pattern, is again a favouring of the top face

over the bottom face (Fig. 3.31, Fig. 3.32, Section B.4[4]). For most of the *Plasmodium* species this distribution is concentrated over one half of the top face, near $\alpha 1$ and the $\beta 7$ and $\beta 1$ C-termini of AdoMetDC. It is worth noting that this concentration of interactions leaves the active site (near the C-terminus of $\beta 3$) exposed for ligand binding. The exception to this is *P. vivax*, which shows most of the top-face covered by ODC COMs (Fig. 3.32). *P. falciparum* shows a slightly greater number of COMs on the bottom face compared to the other species, although this is mirrored slightly by the primate infecting *P. vivax* and *P. knowlesi*. When comparing sides 1 to 4, more variation in the distribution of COMs is observed. The distributions for the rodent parasites (*P. berghei*: Fig. 3.31 and *P. yoelii*: Fig. B.34) show the most similarity. The lack of COMs on the bottom face for rodent parasites compared to the primate parasites is more noticeable in the side views. Within the primate parasites (*P. falciparum*, *P. knowlesi* and *P. vivax*) there are some noticeable differences in the side view distributions. The most noticeable difference occurs in *P. knowlesi* which shows a much greater density of COMs in the vicinity of $\alpha 6$ and $\alpha 7$ (Fig. 3.20) compared to *P. vivax* and in particular *P. falciparum* (side 2 and side 3). In general, the distribution of COMs around *P. falciparum* is more evenly spread.

The concentration of COMs on the top face is in broad agreement with the sequence divergence results, which also show greater sequence conservation for the top face. However, the sequence conservation for region near $\alpha 1$ and the C-termini $\beta 7$ and $\beta 1$ is not particularly well conserved. This does not rule out the possibility of conserved interactions between AdoMetDC and ODC, however.

[4] *P. berghei* is included as representative of four out of the five species, while *P. vivax* is slightly atypical and therefore discussed further here. The remaining results are included in the Appendix (B.4)

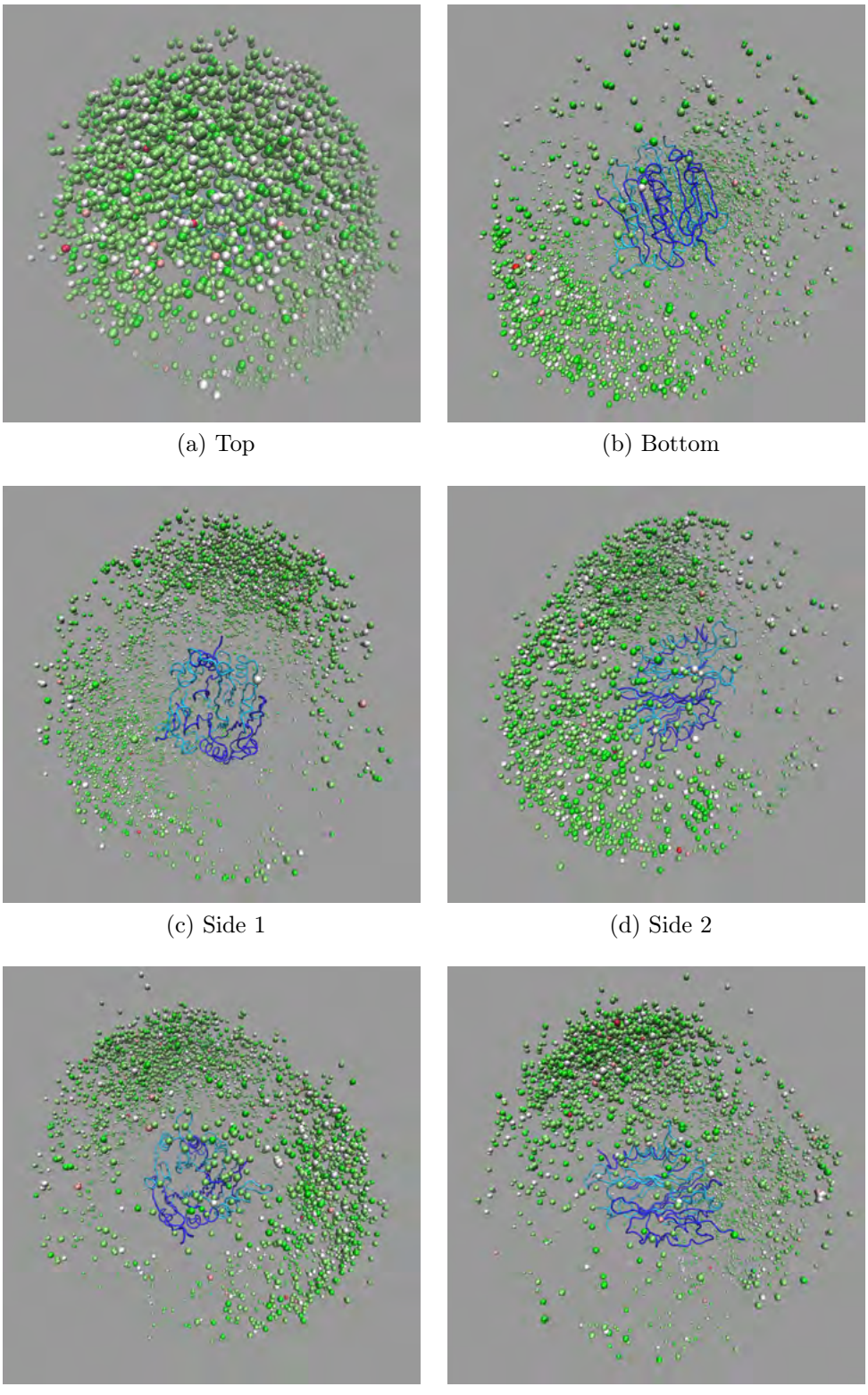


(a) Top (b) Bottom

(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

Figure 3.31: Centre of mass (COM) distributions of *P. berghei* ODC relative to AdoMetDC. AdoMetDC chains A & C (blue) and (cyan) B & D are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based on the RP score (4.73: green → 11.30: red).



(a) Top (b) Bottom

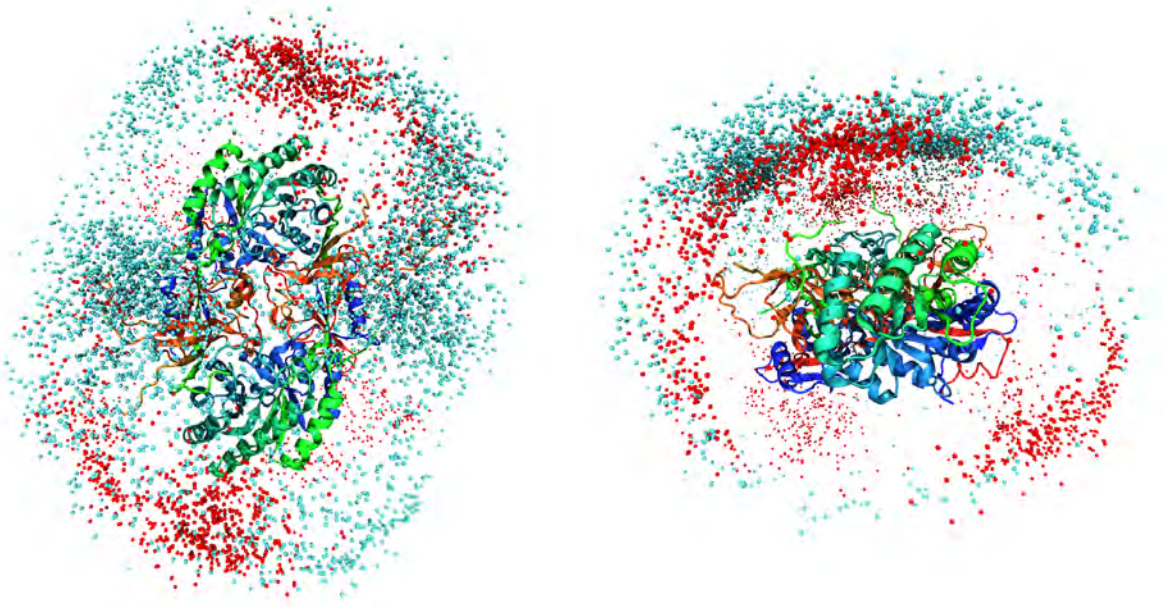
(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

Figure 3.32: Centre of mass (COM) distributions of *P. vivax* ODC relative to AdoMetDC. AdoMetDC chains A & C (blue) and (cyan) B & D are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based on the RP score (4.34: green → 11.42: red).

As a control the same docking protocol was run for all current human structures of Ado-

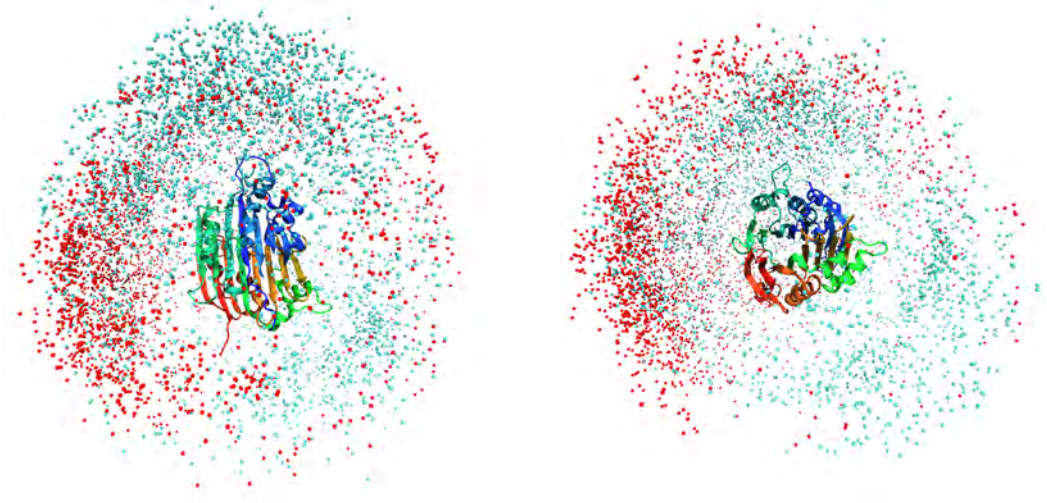
MetDC and ODC for all pairwise combinations. Compared to *Plasmodium,* the distribution of human AdoMetDC around ODC is substantially different. Visual inspection reveals two main distributions for AdoMetDC relative to ODC. In the most prominent of these human AdoMetDC is concentrated around $\alpha 9$, $\alpha 10$, $\alpha 11$ from the TIM-barrel domain of one ODC monomer and the loop between $\beta 12$ and $\beta 13$ in the sheet domain of the other ODC monomer. The second distribution is concentrated around $\beta 1$, $\alpha 1$ and $\alpha 16$ between the TIM-barrel and sheet domains (Fig. 3.33). The distribution of human ODC relative to AdoMetDC is also significantly different compared to the *Plasmodium* models. The human COMs cluster on side 4 near $\alpha 5$, $\beta 7$ and $\beta 15$ (Fig. 3.20), compared to the *P. falciparum* and other *Plasmodium* COMs which cluster near the top face (Fig. 3.34).



(a) Top view of human and *P. falciparum* AdoMetDC COMs

(b) Side view of human and *P. falciparum* AdoMetDC COMs

Figure 3.33: Comparison of COMs of AdoMetDC relative to ODC distributions in human and *P. falciparum*. The distribution of human COMs is substantially different from *P. falciparum* (see main text for further details). Red: human COMs. Cyan: *P. falciparum* COMs.



(a) Top view of human and *P. falciparum* ODC COMs

(b) Side view of human and *P. falciparum* ODC COMs

Figure 3.34: Comparison of the COMs of ODC relative to AdoMetDC distributions in human and *P. falciparum*. The distribution of human COMs is substantially different from *P. falciparum* (see main text for further details). Red: human COMs, cyan: *P. falciparum* COMs.

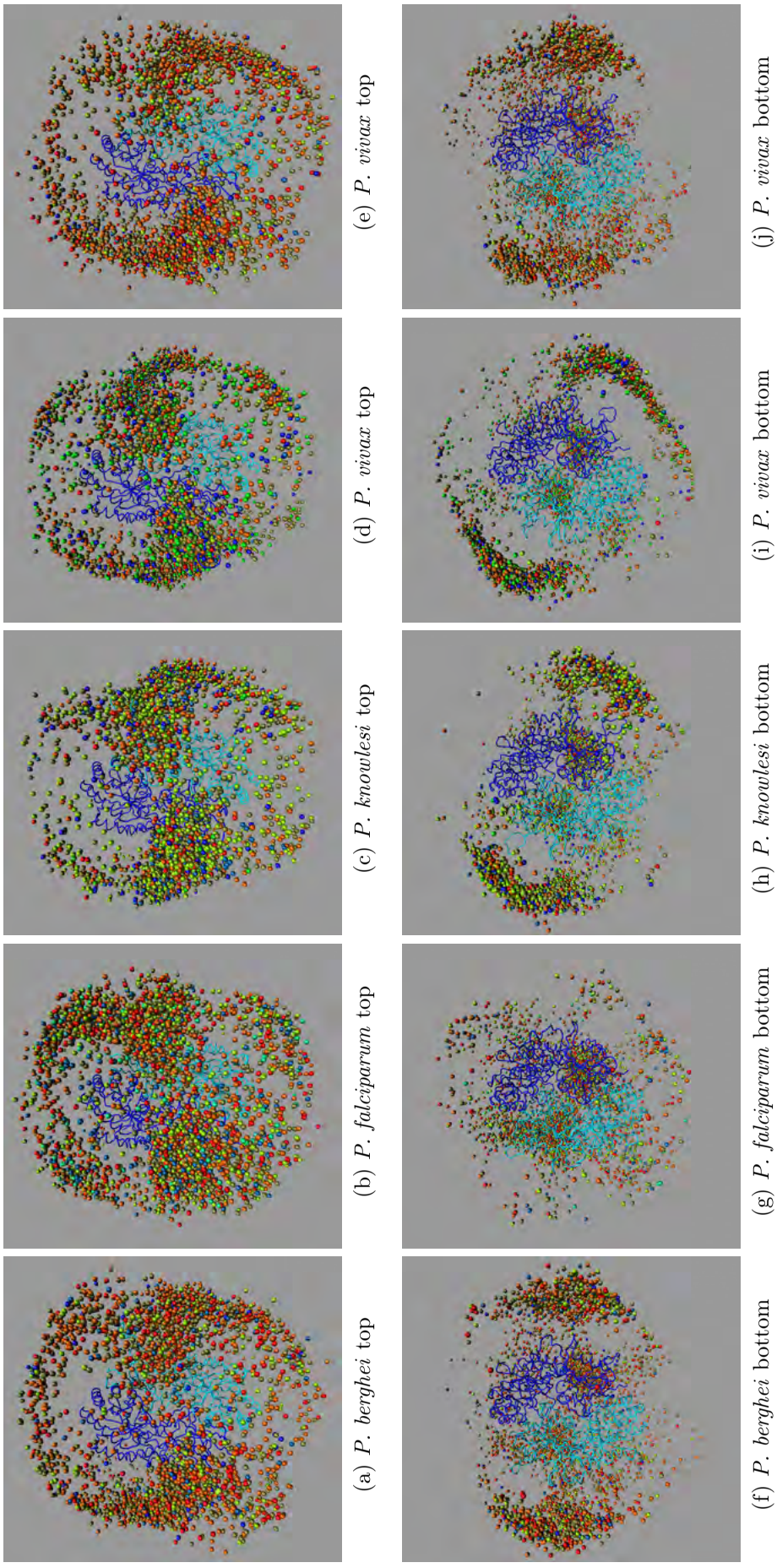
3.4.2.3 Mutually favoured contacting regions

The results of the COM distribution suggest that certain regions on both AdoMetDC and ODC are favoured for making contact within the bifunctional complex. It was further considered that this might be manifested as certain sub-regions on each domain favouring contact with each other over other sub-regions. This was first analysed by determining the closest contacting residue within each docking pose. If certain regions favour contact then residues close in primary sequence should be predicted to be the closest contacting residue for a particular region. This can be visualised by using a colour gradient for residue number (Red → Green → Blue): mutually contacting regions should be visible as patches of similar colour when viewing the COM distribution of one domain relative to another.

Analysis by this means doesn't reveal any obvious mutually contacting regions across the different *Plasmodium* species, however (Fig. 3.35, Fig. 3.36, Fig. 3.37). The most notable difference between species is apparent for the distribution of AdoMetDC COMs relative to ODC in *P. falciparum*. This can be explained by the general absence of contacts made with α1, α16, β2 and α2 compared to the other *Plasmodium* species (Section B.4). A more fine-grained analysis was therefore undertaken to discover conserved contacts.

3.4.2.4 Conserved interactions between AdoMetDC and ODC

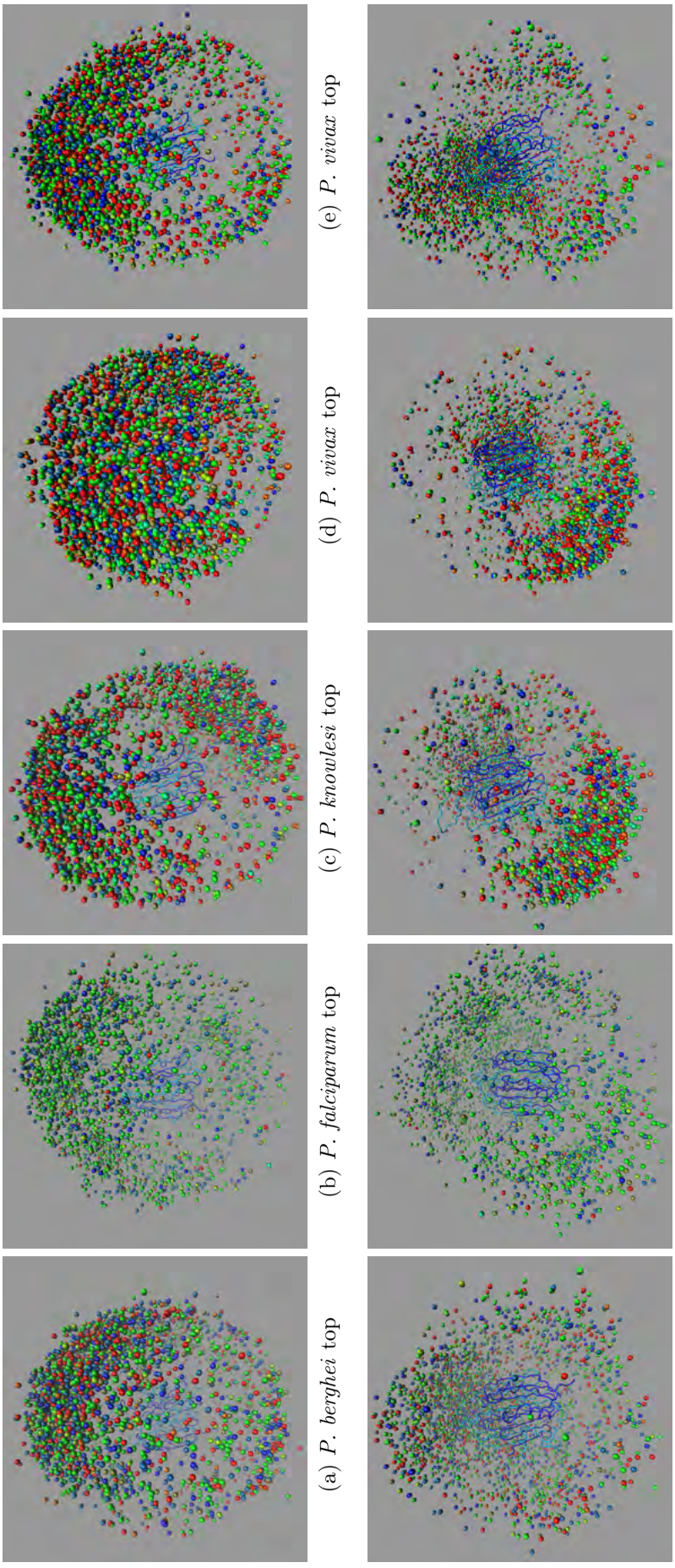
The data visualisation program OPENDX was used to construct heat-maps of the number of times a certain AdoMetDC/ODC residue pair made contact within the 3600 docking runs analysed. Heat-maps were constructed for each *Plasmodium* species. The distance chosen was 15 Å for the C_{α}^{ADC}–C_{α}^{ODC} distance. The distance was chosen based on previous reports



(a) *P. berghei* top (b) *P. falciparum* top (c) *P. knowlesi* top (d) *P. vivax* top (e) *P. vivax* top

(f) *P. berghei* bottom (g) *P. falciparum* bottom (h) *P. knowlesi* bottom (i) *P. vivax* bottom (j) *P. vivax* bottom

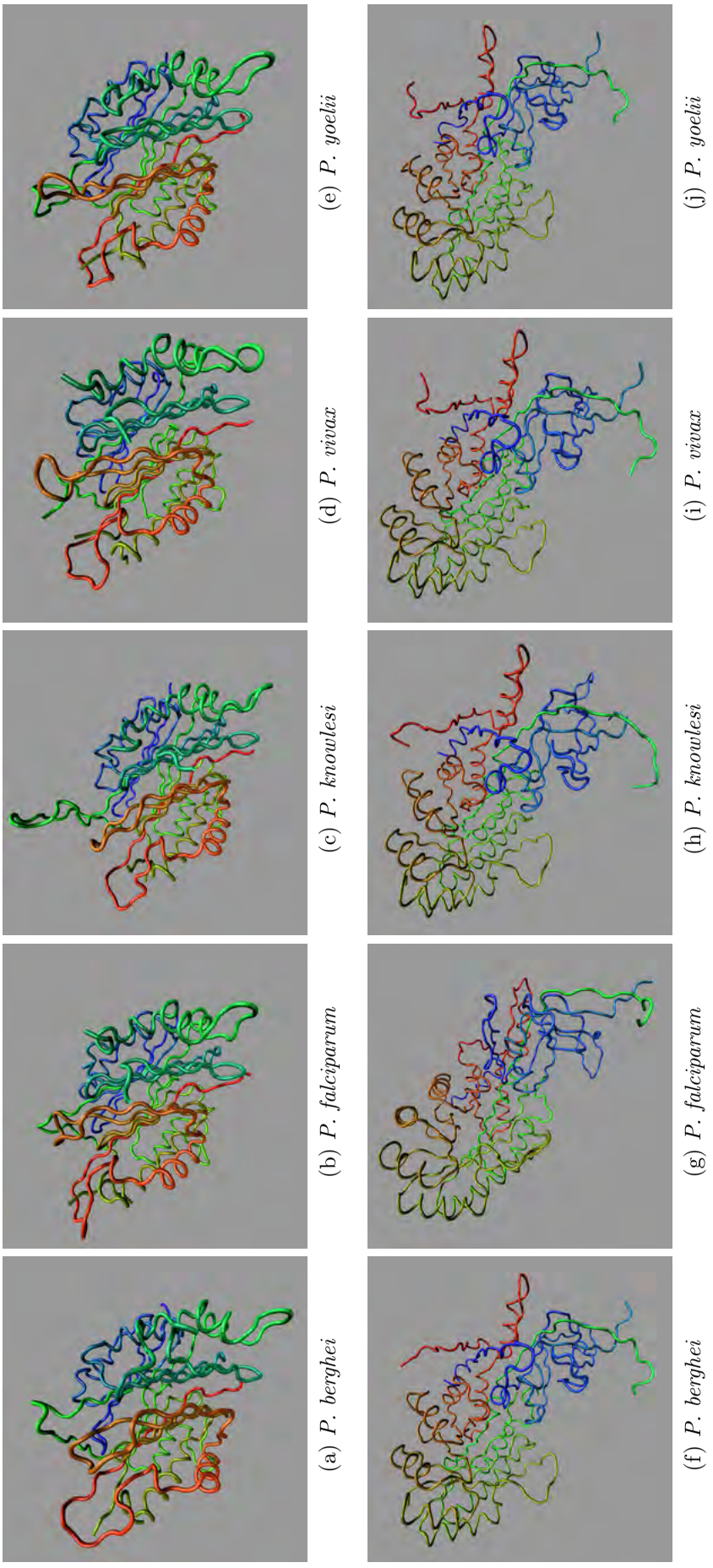
Figure 3.35: AdoMetDC COMs coloured according to closest contacting residue. Red → Green → Blue = increasing residue number (Fig. 3.37). There is no obvious clustering according to colour, i.e. there isn't obvious preference for AdoMetDC/ODC surface patches to associate.



(a) *P. berghei* top (b) *P. falciparum* top (c) *P. knowlesi* top (d) *P. vivax* top (e) *P. vivax* top

(f) *P. berghei* bottom (g) *P. falciparum* bottom (h) *P. knowlesi* bottom (i) *P. vivax* bottom (j) *P. vivax* bottom

Figure 3.36: ODC COMs coloured according to closest contacting residue. Red → Green → Blue = increasing residue number (Fig. 3.37). There is no obvious clustering according to colour, i.e. there isn't obvious preference for AdoMetDC/ODC surface patches to associate.



(a) *P. berghei* (b) *P. falciparum* (c) *P. knowlesi* (d) *P. vivax* (e) *P. yoelii*

(f) *P. berghei* (g) *P. falciparum* (h) *P. knowlesi* (i) *P. vivax* (j) *P. yoelii*

Figure 3.37: AdoMetDC (top row) and ODC (bottom row) coloured by residue position. Red → Green → Blue = increasing residue number. The colour gradient is used to depict the closest residue of each AdoMetDC and ODC docking (Fig. 3.35 and Fig. 3.36).

concerning the standard threshold for two residues to be in contact. According to Punta and Rost (2005) a maximum distance of 8 Å between the C_β atoms implies contacting residues. Adding twice the standard C_α–C_β distance of $\simeq$1.5 Å gives 11 Å. This was increased slightly due to the course grained nature of the methodology to give a threshold of 15 Å.

The heat-maps for all species are generally similar, in that regions that are aligned (according to the alignment used for modeling) are predicted to make contact more often than non-aligned regions. The most notable difference is for the *N*-terminal region of *P. falciparum* ODC where fewer contacts are predicted (Fig. 3.38, Appendix B.5.0.1). This corresponds to the results found for the distribution of COMs, where AdoMetDC was predicted to bind less often to α1, α16, β2 and α2 in *P. falciparum* than in the other species.

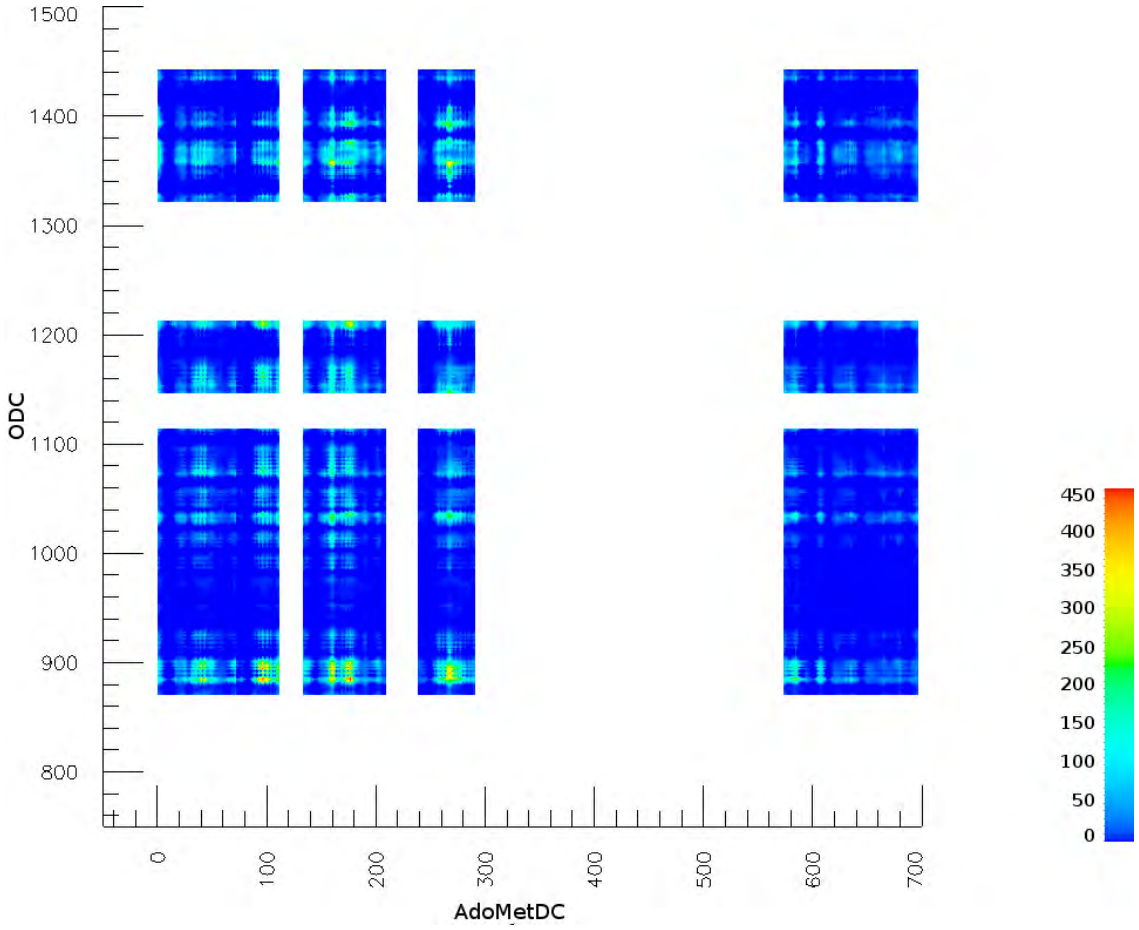


Figure 3.38: Contact count heat-map for *P. berghei*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_α^{ADC}-C_α^{ODC}<15 Å). The typical range is between 0-500, out of a possible 3600.

The heat-maps were restricted to identical residue pairs conserved across all *Plasmodium* species (Fig.3.39). This restricts the number of potential targets for a first round of experiments to test the predicted quaternary structure of AdoMetDC/ODC. From these maps it can be seen that a relatively small number of regions are predicted to make contact often across all five species. The heat-maps were designed to be used interactively, thus allowing a conserved residue pair to be queried for all five species (Fig. 3.40, Appendix B.5.1).

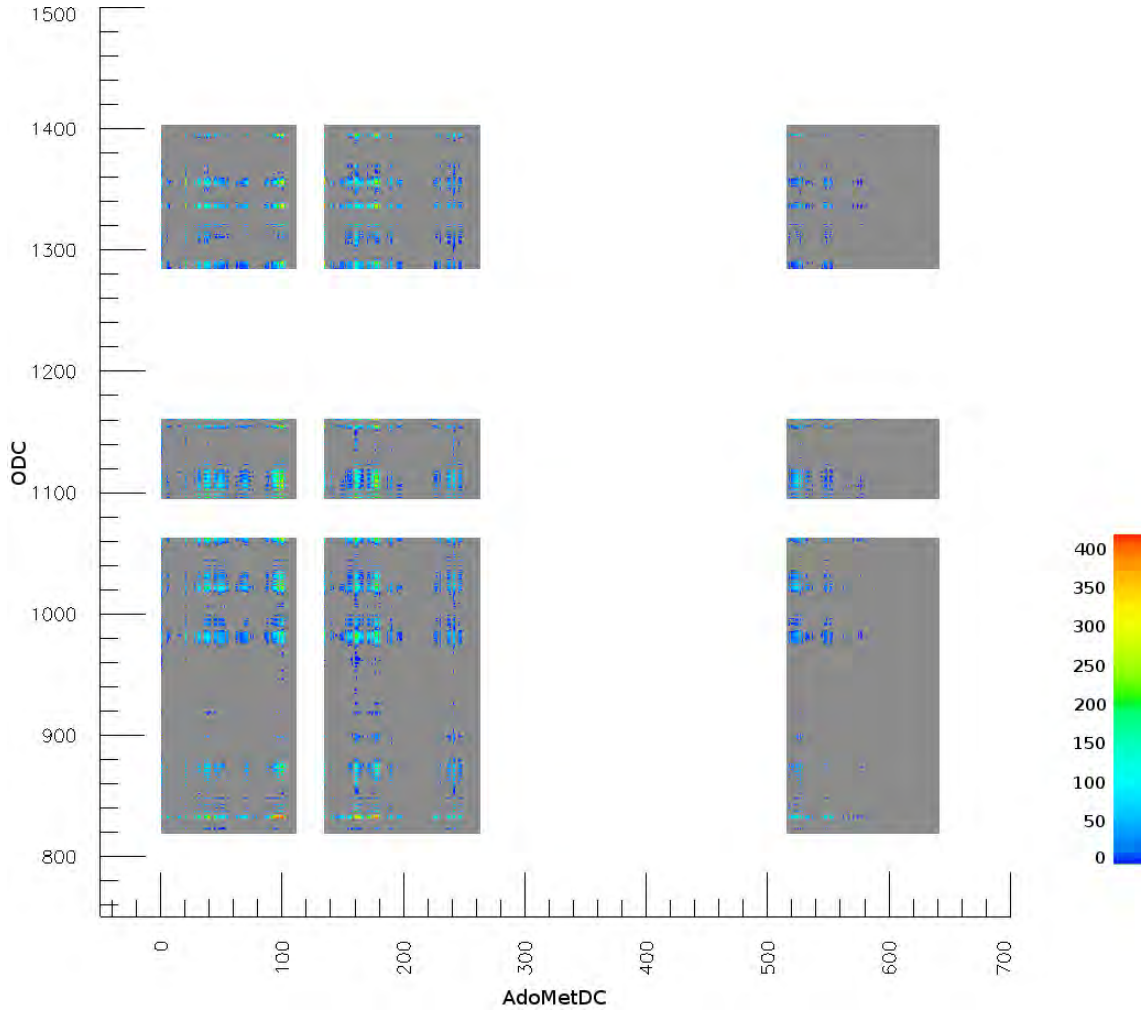
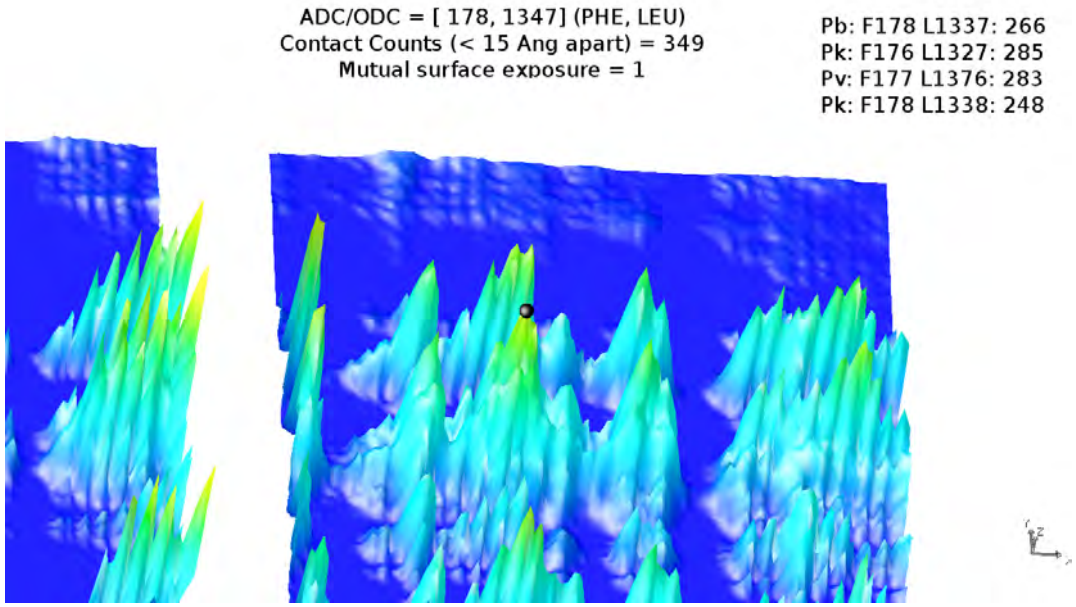
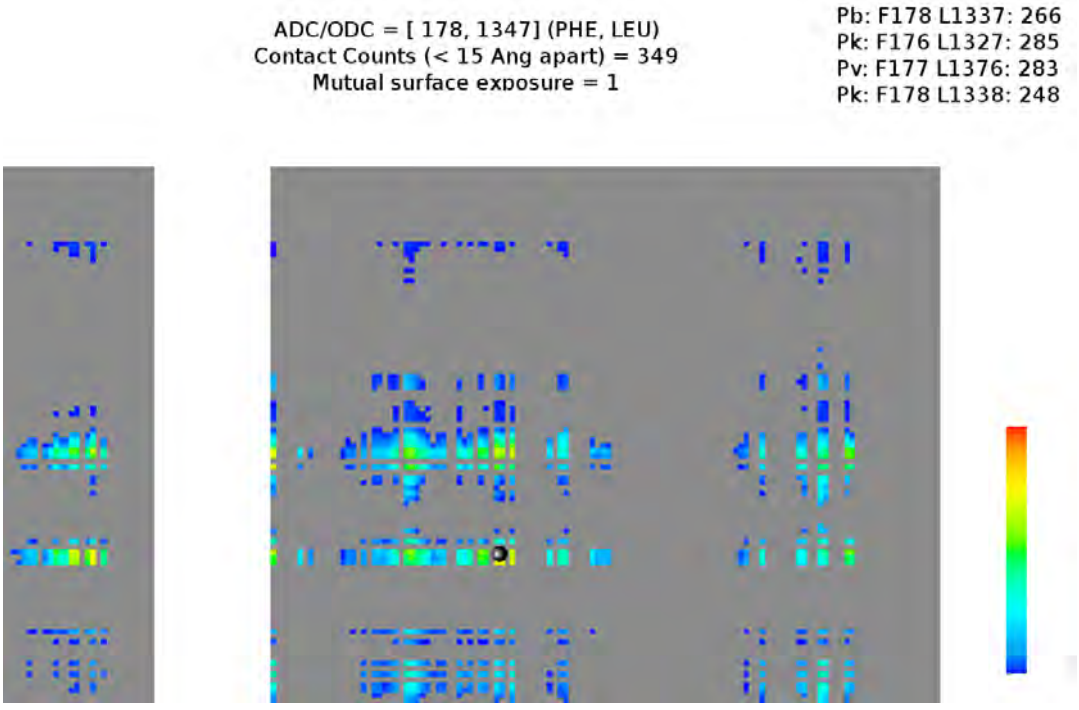


Figure 3.39: *P. berghei* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes contact (C_α^{ADC}-C_α^{ODC}<15 Å). Counts of zero are indicated in light-grey. The typical range is between 0-450, out of a possible 3600.



(a) Warp view



(b) Flat view

Figure 3.40: Zoomed in region of the *P. falciparum* contact count heat-map (identical residue pairs). An AdoMetDC/ODC residue pairing can be selected (black sphere) in either warp or flat view. The contact count number is represented by the blue→green→red colour gradient and height (warp view only). The corresponding contact counts are also displayed for the other *Plasmodium* species.

When comparing all the heat-maps for identical residues there are two regions that show consistently high counts across all five species (Fig. 3.41). These regions correspond to one and two contiguous surface patches on AdoMetDC and ODC, respectively.

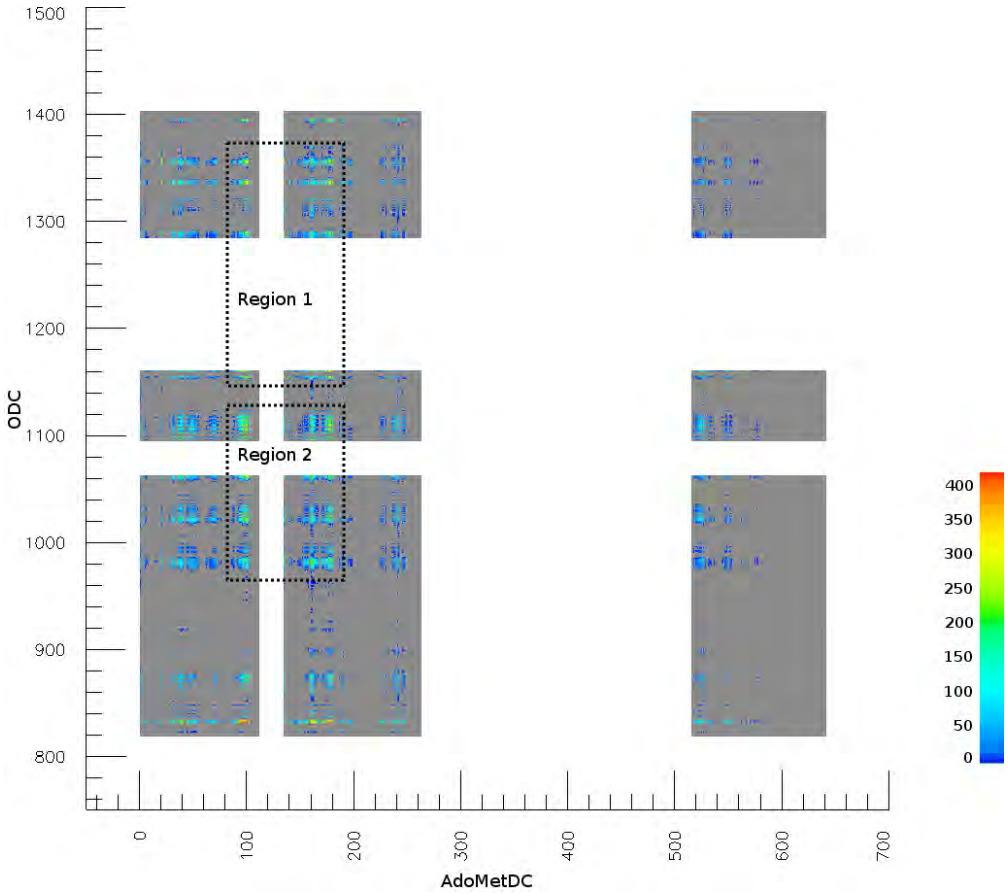
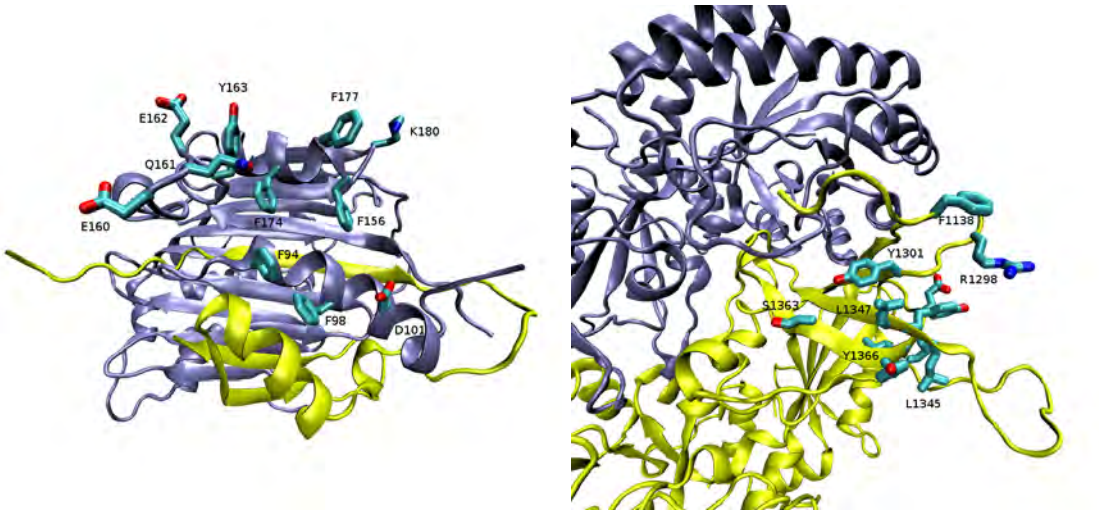


Figure 3.41: Region 1 and 2 of conserved contacting regions between AdoMetDC and ODC. See text for further details.

Region 1 Region 1 represents two patches dominated in the centre by conserved hydrophobic residues, with conserved hydrophilic and charged residues on the rim. Both ODC and AdoMetDC patches are found on the top face of their respective proteins (Fig. 3.42). This is in agreement with COM distribution and sequence conservation. The ODC patch forms part of the surface of the sheet domain. The numbering used is as for *P. falciparum*. The hydrophobic core of the AdoMetDC patch includes Tyr 163, Phe 174, Phe 177, Phe 178, Phe 94, and Phe 98. The rim of the AdoMetDC patch includes Glu 101, Glu 160, Gln 161, Glu 162 and Lys 180.

The hydrophobic core of the ODC patch includes Phe 1138, Leu 1345, Leu 1347, Tyr 1301. The rim of the ODC patch includes Arg 1298 and Ser 1363. The ODC patch contains further conserved residues whose sidechains form part of the protein core, not the surface, and are therefore not listed. The core of this region shows the highest contact numbers for the residues listed in Table 3.3. Furthermore, the high number of contacts between between Glu 160 and Arg 1298 and Glu 162 and Arg 1298 suggest that both these pairs may be involved in potential salt-bridge interactions. A further salt-bridge is possible between Arg 198 and Asp 101.

The ODC patch is also predicted to contain the parasite-specific insert O_2 between Phe 1138 and Arg 1298. Insert A_2 of AdoMetDC is also predicted to be near the AdoMetDC patch, although not within it.



(a) AdoMetDC (b) ODC

Figure 3.42: Conserved residues involved in the region 1 AdoMetDC/ODC interaction (numbering as for *P. falciparum*). For ODC only those conserved residues whose sidechains are oriented away from the surface are listed. See text for further details.

Region 2 Region 2 comprises the same AdoMetDC patch involved in region 1 and a conserved patch on the top-face of the ODC TIM-barrel domain. Unlike the ODC patch for region 1, hydrophobic and hydrophilic residues are interspersed. The ODC patch comprises Phe 960, Lys 961, Tyr 963, Lys 964, Ser 965, Met 967, Lys 1004, Asn 1005, Cys 1010, Lys 1014, Leu 1015, Arg 1017, Tyr 1085, Glu 1088, Lys 1089, Leu 1092, Ala 1093, Met 1096, Ser 1097, His 1100 and Tyr 1101. The ODC patch is predicted to form part of the surface of the TIM barrel domain and corresponds to the second distribution of COMs described above making contact near α1, α16, β2 and α2. Compared to region 1 the contact count is generally less by 50 to 100.

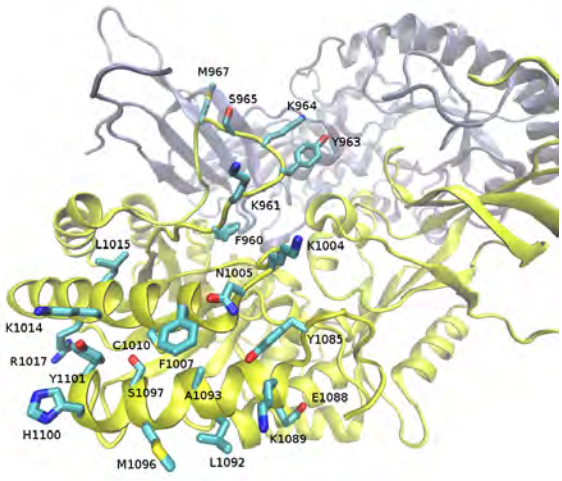


Figure 3.43: ODC patch of region 2. Numbering as for *P. falciparum*. For ODC only those conserved residues whose sidechains are oriented away from the surface are listed.

Table 3.3: Core hydrophobic residues and potential salt-bridge contacts in region 1. cc = "contact count", i.e. the number of dockings for which the C_{α}^{ADC}-C_{α}^{ODC} distance is ≤15 Å out of a total of 3600.

P. berghei			*P. falciparum*			*P. knowlesi*			*P. vivax*			*P. yoelii*		
AdoMetDC	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*
F178	L1337	266	F178	L1347	349	F176	L1327	285	F177	L1376	283	F178	L1338	248
F178	Y1336	259	F178	Y1346	325	F176	Y1326	239	F177	Y1375	235	F178	Y1337	235
F178	L1335	281	F178	F1345	330	F176	F1325	186	F177	F1374	190	F178	F1336	223
F177	L1337	250	F177	L1347	351	F175	L1327	306	F176	L1376	297	F177	L1338	217
F177	Y1336	249	F177	Y1346	336	F175	Y1326	269	F176	Y1375	268	F177	Y1337	206
F177	F1335	262	F177	F1346	340	F175	F1325	217	F176	F1374	249	F177	F1336	233
K180	L1337	232	K180	L1347	312	K179	L1327	263	K180	L1376	229	K180	L1338	221
K180	Y1336	241	K180	Y1346	316	K179	Y1326	236	K180	Y1375	230	K180	Y1337	228
K180	F1335	249	K180	F1346	327	K179	F1325	206	K180	F1374	230	K180	F1336	241
E160	R1287	137	E160	R1298	167	E158	R1275	114	E159	R1324	157	E160	R1288	105
E162	R1287	140	E162	R1298	151	E160	R1275	96	E161	R1324	167	E162	R1288	88
D100	R1287	182	D101	R1298	219	D99	R1275	84	D100	R1324	76	D100	R1288	151

Table 3.4: Core hydrophobic residues and potential salt-bridge contacts in region 2. cc = "contact count", i.e. the number of dockings for which the C_{α}^{ADC}-C_{α}^{ODC} distance is ≤15 Å out of a total of 3600.

P. berghei			*P. falciparum*			*P. knowlesi*			*P. vivax*			*P. yoelii*		
AdoMetDC	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*	*AdoMetDC*	*ODC*	*cc*
E160	K978	180	E160	K961	204	E158	K974	196	E159	K1030	202	E160	K978	178
E162	K978	124	E162	K961	110	E160	K974	145	E161	K1030	153	E162	K978	110
E160	K981	216	E160	K964	288	E158	K977	221	E159	K1033	237	E160	K981	186
E162	K981	195	E162	K964	226	E160	K977	192	E161	K1033	225	E162	K981	152
E160	K1021	178	E160	K1004	168	E158	K1017	182	E159	K1073	205	E160	K1021	202
E162	K1021	163	E162	K1004	147	E160	K1017	171	E161	K1073	155	E162	K1021	159
E160	K1031	86	E160	K1014	118	E158	K1027	91	E159	K1083	137	E160	K1031	129
E162	K1031	74	E162	K1014	109	E160	K1027	82	E161	K1083	156	E162	K1031	104
E160	R1034	64	E160	R1017	99	E158	R1030	64	E159	R1086	110	E160	R1034	117
E162	R1034	53	E162	R1017	80	E160	R1030	52	E161	R1086	112	E162	R1034	80
E160	K1031	86	E160	K1014	118	E158	K1027	91	E159	K1083	137	E160	K1031	129
E162	K1031	97	E162	K1014	134	E160	K1027	86	E161	K1083	155	E162	K1031	141
E160	K1106	131	E160	K1089	161	E158	K1102	120	E159	K1158	150	E160	K1106	250
E162	K1106	114	E162	K1089	134	E160	K1102	114	E161	K1158	136	E162	K1106	212
E160	H1117	166	E160	H1100	158	E158	H1113	83	E159	H1169	142	E160	H1117	230
E162	H1117	135	E162	H1100	148	E160	H1113	77	E161	H1169	142	E162	H1117	205

3.5 Conclusion

The bifunctional arrangement of *Plasmodium* AdoMetDC/ODC is unusual as the two functions appear to be in separately expressed proteins in all other organisms. Previous studies suggest that interfering with the formation of the bifunctional complex may serve as parasite-specific route to inhibition. This requires detailed knowledge of the quaternary structure of the complex, however. Due to the absence of experimental structures protein-protein docking of homology models of both domains was performed in order to inform the design of further experiments to probe the quaternary nature of the complex. By sampling a number of models with different conformations across multiple *Plasmodium* species it was possible to overcome the uncertainties inherent in using homology models for predicting protein-protein interactions. From this course-grained first pass experiments can be designed to test the predictions and allow follow-up with more refined docking.

The results of the above experiments suggest that for each domain one face each of both AdoMetDC and ODC is likely to be favoured for making contact between the domains within the bifunctional complex. In each domain, the arbitrarily designated "top" face is favoured for contact. This is first suggested by sequence conservation within the five *Plasmodium* AdoMetDC/ODC sequences studied. There is greater conservation on both the "top" faces of AdoMetDC and ODC according to the alignments used for modeling, compared to the rest of the domain surfaces. This is confirmed by docking of AdoMetDC against ODC. In order to minimise spurious results, multiple models of each domain were docked against each other within each *Plasmodium* species. The conformation of each model is slightly different, thus allowing a wider range of possible dockings to be explored and minimising the chance of high scoring predicted contacts being artifacts of the docking process. From the various pairwise docking runs the top 100 scoring poses were analysed based on the residue-pairwise based RP Score of FTDOCK, thus for each species 3600 poses where chosen. The centre of mass distribution (COM) for each domain and it's docking partner within each species further suggests that top faces are favoured for contact. Over the top face of ODC there is a greater distribution around the sheet domain which is conserved across species. This would also bring AdoMetDC within contact of the inserts O_1 and O_2. This confirms previous results that deletion of these inserts affects AdoMetDC activity (Birkholtz *et al.*, 2004), which would be more likely if O_1 and O_2 were in direct contact with AdoMetDC. Finer grained analysis of which AdoMetDC/ODC residue pairs make contact more often provides further insight. Two patches were identified on the surface of ODC, while one was found on the surface of the smaller AdoMetDC domain. The ODC patch of region 1 corresponds to part of the sheet domain, also predicted to contain the insertion site of O_2. This patch is dominated by hydrophobic residues which corresponds with previous studies where protein-protein interactions largely comprise the burial of mutual hydrophobic patches (Young *et al.*, 1994; Lijnzaad *et al.*, 1996). The corresponding AdoMetDC patch is also fairly hydrophobic. Region 1 also contains residue pairs that may form salt-bridges, and are predicted to be good candidates for further mutagenesis studies. The ODC patch of region 2 is predicted

to be a less likely contact region for AdoMetDC for a number of reasons. Firstly, the COM distribution is lower than for region 1. Secondly, the distribution of hydrophobic and hydrophilic residues is more mixed compared to region 1. Lastly, AdoMetDC is known to form a dimer in both *Plasmodium* and mammals. If AdoMetDC were to make contact with ODC in region 2, the two AdoMetDCs would not be close enough to also make direct contact. Thus region 1 is considered the best place to target AdoMetDC/ODC complex formation.

A number of caveats in this study also need to be mentioned. Firstly, the long inserts of A_2, A_3 and O_1, O_2 were left un-modelled due to the uncertainty involved in modeling such long stretches *ab initio*. These inserts may well be involved in complex formation, however. This has already been demonstrated for O_1, where deletion prevents formation of AdoMetDC/ODC hybrid complexes with separately expressed domains (Birkholtz *et al.*, 2004). Secondly, while protein-protein interfaces typically involved burial of hydrophobic patches, the predicted interfaces differ from known interactions in some important respects. Protein-protein interactions are often characterised by the presence of so-called hot-spots, usually defined as residues where mutation to Ala increases the free energy of binding by at least 2.0 $kcal/mol$ (Thorn and Bogan, 2001). Hot-spots show a specific residue bias, and tend to be enriched by certain residue such as Trp, Tyr and Arg. Furthermore, some hydrophobic residues which are present in interacting patches are still less favoured in hot-spots. Notably, Ile is favoured over isomeric Leu (Bogan and Thorn, 1998). While rich in hydrophobic residues, the patches of regions 1 and 2 are not enriched in Trp, Tyr and Arg. Within protein-protein interacting sites hot-spots and polar residues are posited to be surrounded by a ring of hydrophobic residues. This so-called "O-ring" serves to exclude solvent from the polar residues thus strengthening their interactions. The O-ring structure is not evident in regions 1 and 2, however the importance of O-rings in protein-protein interactions remains inconclusive (Moreira *et al.*, 2007). In the *Plasmodium* models there are no conserved surface Trp and Arg residues within regions 1 and 2 (results not shown). While there are conserved Tyr residues, they don't appear to be surrounded by an O-ring. Therefore, according to the alignments used hot-spots as currently understood, do not occur on the surface of *Plasmodium* AdoMetDC or ODC.

In this study only the docking of the AdoMetDC monomer with ODC dimer was investigated. The ODC dimer could be reliably predicted by homology modeling with dimeric templates. While *P. falciparum* AdoMetDC is known to interact (Birkholtz *et al.*, 2004), modeling of the AdoMetDC dimer was not attempted in this study. According to the dimeric structure for the mammalian cognate, *P. falciparum* AdoMetDC is predicted to interact via association of $\beta 7$ and $\beta 15$, which corresponds to side 4. The sequence divergence of this region is higher compared to the rest of the structure, and finding an alignment proved difficult. It is considered highly likely that protein-protein docking with respect to this region will yield unreliable results. Furthermore, the interaction surface area would be small. The software used was found not to perform well with re-docking of the ODC template dimers during attempts to reproduce the known structure (results not shown). However, human

ODC is also known to contain a relatively small surface area compared to other protein complexes (Kern *et al.*, 1999). Therefore, for determining AdoMetDC-AdoMetDC contacts it is likely to be more fruitful to extrapolate from AdoMetDC ODC-dimer orientations that allow for AdoMetDC-AdoMetDC to interact. Due to the course grained nature of this study it is considered that this should await the first round of experimental testing.

The contact count suggests that salt-bridges may be possible within region 1 and 2, but they are predicted to be on the periphery and not within a hydrophobic patch. Hydrogen-bonding and salt-bridges within protein interfaces tend to be highly buried in order to limit interactions with solvent (Xu *et al.*, 1997). However, due to the uncertainty regarding the structure of the hinge and inserts, it cannot by ruled out that the salt-bridges may still be protected from solvent by these elements.

The results of the docking experiments can be used to design a first pass of experiments to probe the quaternary structure of *Plasmodium* AdoMetDC/ODC. Specifically the conserved residue pairs of region 1 and 2 with high contact counts should be targeted by site-directed mutagenesis. Alanine substitution ("alanine-shaving") followed by determining the free energy of association is a common technique used for this (Bogan and Thorn, 1998).

The computational predictions performed are a first pass and should be considered course grained. Experimental confirmation of the interacting surfaces would justify more fine-grained computational approaches. Future computational studies could include:

- Allowing for side-chain movement of the interacting surfaces
- Docking of full-atom models
- Prediction of AdoMetDC-AdoMetDC contacts within the bifunctional complex
- *Ab initio* modeling of inserts
- Comparing results from different docking programs (e.g. ROSETTA, Z-DOCK)

Results from such studies could then be used to direct the next round of experiments. Firstly site-directed mutagenesis may be used to confirm interacting regions, especially via careful measurement of energetic changes using methods such as iso-thermal-calorimetry. This could be followed by synthesis of probes designed to disturb the AdoMetDC/ODC interfaces as a means to disrupt enzyme function. Furthermore, currently wet-lab experiments are performed on recombinantly expressed AdoMetDC/ODC from *P. falciparum* only, whereas the computational experiments include all five *Plasmodium* species for which the full bifunctional sequence is known. The quaternary structure is expected to be conserved across all species, however. Therefore, a more thorough experimental approach would involve repeating the confirmation of predicted interactions for recombinantly expressed proteins from all five species. If domain-only crystal structures are obtained instead of the full-length protein, these should be used for docking instead of homology models. Through an iterative

cycle and experimentation it should therefore be possible to determine the quaternary structure of *Plasmodium* AdoMetDC/ODC and assess non-active site based targets within the bifunctional complex.

Chapter 4

Concluding Discussion

Malaria remains a significant world health problem, potentially affecting hundreds of millions of people, particularly in developing countries. The threat is likely to increase due to the increasing spread of resistance to currently available drugs and the onset of global warming. A substantial improvement in the economic prospects of the developing world will require reliable means of preventing and treating this debilitating disease. For this, the timeous discovery and deployment of new anti-malarials will be essential (Section 1.2). Historically drug discovery has relied heavily on random screening of natural and man-made chemicals for therapeutic activity. This has lead to the development of high-throughput screening techniques combined with combinatorial methods to generate large libraries of compounds numbering in the thousands to millions. It is now routinely possible to screen these large libraries for activity at the molecular or cellular level. However, despite these developments the rate of new drugs reaching the market has actually slowed. While this is partly due the implementation of more stringent safety standards, it is obvious that a blind throughput approach will not explore chemical space quickly enough to discover new active drugs. Much of the high attrition rate of new drug leads is due to failures late in the drug discovery process involving bio-availability and toxicity. Therefore, various methods have recently emerged that aim to restrict the search of chemical space to those regions most likely to show activity at the molecular level and at the same time less likely to fail at the higher physiological levels. These methods fly under the large banner of "Rational Drug Discovery" and are largely computational in nature. Methods range from optimising ligand-protein binding with atomistic modeling and large scale *in silico* screening of compound libraries to prediction of drug-likeness properties associated with good ADMET characteristics and ease of synthesis (Section 1.3).

This study focused on the application of methods at the atomistic and molecular level to three proteins from the polyamine biosynthesis pathway of the malaria parasite. The application of these methods depends on detailed structural knowledge of the proteins in question. While the most reliable structures are derived from experimental methods such as X-ray crystallography and NMR, the rate of structure determination lags significantly behind other high-throughput biological methods such as sequence determination, array ex-

periments and large-scale imaging. This is largely due to the resource intensive nature of current methods and that only a small number of proteins prove amenable to them. New techniques under development promise to allow the determination of individual macromolecule structures in solution (Mardis *et al.*, 2009; Tiede *et al.*, 2009), thus allowing any target to be investigated. However, these are unlikely to become routine in the short term.

In the absence of experimentally determined structures computational methods can be used to fill the gap. These can range from *ab initio* structure prediction to the most common method used, so-called knowledge based or "homology modeling", whereby proteins with a common evolutionary ancestor can be modelled if the structures of one or more its relatives are known. This is particularly useful in *Plasmodium* where experimental determination of protein structures has proved particularly refractory. Recombinant expression of *Plasmodium* proteins in the standard *E. coli* system has traditionally proved difficult due parasite-specific characteristics of *Plasmodium* proteins. Among these are the high non-*E. coli*-like A+T rich content of *Plasmodium* genes (70-80%) and the larger average size of typical *Plasmodium* proteins. The larger size is usually manifested by *Plasmodium*-specific inserts within the core structure relative to other homologues. These inserts are often characterised by low-complexity and dominated by hydrophilic residues and further compromise expression and crystallisation. Structural modeling is thus particularly useful in the case of *Plasmodium* (Section 1.4).

The characteristics of *Plasmodium* proteins that compromise experimental structure determination also hinder modeling. Specifically, low sequence similarity combined with the presence of inserts make finding a reliable sequence alignment difficult. Nonetheless, homology modeling can be performed by following the careful application of methods that maximise the use of available information. This study describes a number of methods that can be followed to overcome these difficulties. Furthermore, computational methods can be used to gain a further understanding of these *Plasmodium*-specific features. This is particularly useful as these features may themselves provide novel parasite-specific targets for drug inhibition. Among these features are potentially unique protein-protein interactions compared to the host that may serve as novel non-active site targets.

In this study, the application of computational methods to further understand the structure of three proteins from the polyamine biosynthesis pathway of *Plasmodium* is described. Attention is focused on *Plasmodium*-specific features that distinguish these enzymes from their homologues, particularly the human host cognates. For the first enzyme, arginase, homology modeling and molecular dynamics were performed to gain further insights into the structural metal Mn^{2+} dependency of *Plasmodium* arginase. Via molecular dynamics it was possible to further confirm the dependence of arginase trimerisation on the presence of the active site metals (Section 2.4.3). Furthermore, a novel inter-monomer salt-bridge was predicted and further confirmed experimentally via site-directed mutagenesis. Further simulation of the mutant proteins again confirmed the structural metal dependency and sensitivity of the complex to the presence of inter-monomer interactions and active site metal.

The sensitivity to the presence of the metal was robust across different simulation protocols and could be used in future for computational screening of compounds designed to disturb the enzyme via disturbing protein-protein interactions within the trimer. For the second and third enzymes a slightly different approach was followed. Instead of molecular dynamics, protein-protein docking was employed. In the case of arginase and the ODC dimer the conserved interactions could be reliably predicted via modeling on oligomeric templates. Due to the unique bifunctional nature of *Plasmodium* AdoMetDC/ODC this approach could not be followed for predicting the interactions of the domains. The domains of the bifunctional AdoMetDC/ODC enzyme were individually modelled and their interactions predicted using an exhaustive sampling and scoring based approach. From the large number of dockings performed it was possible to predict the most likely regions where AdoMetDC and ODC interact (Section 3.4.2). The predicted regions are auspicious in that they are located close to protein-inserts that are known to have cross-domain activity. These predictions can be used to design further experiments for probing AdoMetDC/ODC interactions, the results of which can be used to determine finer grained docking and analysis.

For all the proteins targeted in this study a number of methods were followed to overcome difficulties typically associated with *Plasmodium* proteins. Obtaining a reliable alignment was aided in all cases by including the sequences of multiple *Plasmodium* species, specifically to increase confidence in the insert boundaries. All protein structural models contain errors, whether they be experimental or computational. The error rate for homology models tends to be higher, however. Correcting all the errors of a model is often not practical, nor necessary for all follow-up applications. Instead of trying to optimise one model, multiple homology models were constructed and selected for further analysis based on various quality checks. For arginase, only one model was chosen since the lengthy nature of the molecular dynamics simulations employed prohibited repeating the experiment with multiple models. For the docking of AdoMetDC and ODC multiple models were docked against each other in an attempt to mask errors that would have been introduced into a particular single model. This process was repeated for five *Plasmodium* species to further increase sampling. For similar simulation time reasons as described above, only the arginase of *P. falciparum* was modelled.

The inclusion of multiple *Plasmodium* species proved particularly useful, and it is suggested that this approach should be followed more thoroughly in the experimental setting as well. This has been partially described for some large scale expression and structural genomics programs involving *Plasmodium*. Specifically, a repeat of studies across all species for arginase and AdoMetDC/ODC could prove insightful. It remains unknown whether the structural metal dependency is unique to *P. falciparum* or whether it occurs in *Plasmodium* in general. Targeting arginase via disturbing trimerisation might therefore prove a more generic strategy. Similar comments follow for AdoMetDC/ODC. To date, all experimental work has focused on *P. falciparum*, and it remains unknown whether the peculiarities occur across all species. This is again important for targeting unique features of AdoMetDC/ODC

as a drug target, but if results are reproducible across species it will increase confidence in current interpretations of existing experimental data.

This study has successfully demonstrated that computational methods can be used to further the understanding of features unique to *Plasmodium* proteins. This was enabled by access to high performance computing resources, whereby thousands of processors can be used in parallel to solve a problem. The results were in agreement with previous experiments and in the case of arginase confirmed by further experimental follow-up. Further application of such approaches can supplement the dearth of structural data. In particular it is becoming possible to simulate large systems numbering in the millions and perform microsecond length simulations of small systems (Dror *et al.*, 2009; Klepeis *et al.*, 2009). This will make the *ab initio* determination of the insert regions feasible, combined either with homology models or experimental data for the protein core. The paucity of structure data is particularly poignant for protein-protein interactions since the number of protein complex structures is significantly smaller. Due to the combinatorial nature of potential interacting proteins, experimental follow-up of each possible combination will be impractical. Therefore, computational methods are likely to become extremely useful in predicting protein-protein interactions. This is yet another example of how computational methods can be used for virtual screening of experiments most likely to be fruitful. Molecular modeling also demonstrates its value in being able to answer questions that current experimental methods cannot. Molecular dynamics is particularly adept at this, allowing visualisation of mechanisms at an atomic level. This was demonstrated using the arginase model to further understand the relationship between metal-binding and oligomerisation. Structure determination and high-throughput methods are likely to improve to the point of rendering certain current computational methods unnecessary. It is not unlikely that it will become possible to determine the structure of any desired protein and not be constrained by the fickle nature of proteins regarding crystallizability. However, it will always remain impractical to perform all possible experiments. Simulation can help bypass this limit.

Experiments can be considered "live simulations" and are better in that the answer is produced quickly. However, in an experiment preparing the system can be time-consuming and difficult. The fundamental drawback of typical experiments in molecular biology involves extracting the answer which often requires amplification of the output to a human visible level. Computer simulations, while often much more time consuming in producing the output, are better in that it is easy to alter the initial conditions and visualising the output doesn't depend on a method to amplify the result. This ability to explore the gaps left by experiment will ensure that Simulation, alongside Theory and Experiment, will become the Third Pillar of Science.