UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Molecular modeling elucidates parasite-specific features of polyamine pathway enzymes of *Plasmodium falciparum*

by

Gordon Andreas Wells

Submitted in partial fulfilment of the requirements for the degree

Philosophiæ Doctor

In the Faculty of Natural & Agricultural Sciences

Department of Biochemistry

University of Pretoria

Pretoria

March 2010

© University of Pretoria

The first principle is that you must not fool yourself – and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that.

– Richard Feynman

Declaration

I, Gordon Andreas Wells, declare that the thesis/dissertation, which I hereby submit for the degree *Philosophiae Doctor* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:..... DATE:.....

Abbreviations and Nomenclature

ADMET	Absorption, Distribution, Metabolism, Excretation and Toxicity
AdoMetDC	*S*-adenosylmethionine decarboxylase
Ama x_a::Ama y_b	Salt-bridge between Ama x of monomer a and Ama y of monomer b
Ama *xyz*	Amino acid *position*
αn	Alpha/3_{10} helix n
βn	Beta strand n
ATP	Adenosine tri-phosphate
AZ	Antizyme
Å	Ångström
BMIC	Bio-Medical Informatics Centre
CAPRI	Critical Assesment of PRedicted Interactions
CHPC	Centre for High Perfomance Computing, Cape Town, SA
CJ	Conjugate gradients
COM	Centre of Mass
CSIR	Council for Scientific and Industrial Research, SA
DDT	Dichlorodiphenyltrichloroethane
DFMO	*alpha*-Difluoromethylornithine
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
EC	Enzyme Commission
EGEE	Enabling Grids for E-sciencE

eIF5A	Eukaryotic initiation factor 5A
IRS	Indoor residual spraying
ITN	Instectiside treated nets
MD	Molecular dynamics
NAP	Nuclear Aggregates of Polyamines
NMR	Nuclear Magnetic Resonance
ODC	Ornithine Decarboxylase
PAO	Polyamine oxidase
PDB	Protein Data Bank
Pf Arg	*Plasmodium falciparum* arginase
*Pf*AdoMetdcDC	*Plasmodium falciparum* AdoMetDC
*Pf*ODC	*Plasmodium falciparum* ODC
PLP	Pyridoxal 5'-phosphate
QSAR	Quantitative Structure Activity Relationship
RMSD	Root Mean Square Deviation
RP	Residue Pairing Potential
SSAT	Spermine/spermidine N^1-acetyltransferase
TIM	Triose-phosphate isomerase
TIP3	Transferable Intermolecular Potential

Contents

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

List of Figures

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

List of Tables

Acknowledgements

My supervisor Prof. A.I. Louw and co-supervisors Prof. F. Joubert, Prof. L-M. Birkholtz and Prof. R.D. Walter for their patience and advice.

The South African National Research Foundation (NRF)

The South African National Bioinformatics Network (NBN)

The NRF German-South African Co-operation Grant

Temporary access was granted to the iQudu cluster at the Centre for High Performance Computing (CHPC), Meraka Institute, Council for Scientific and Industrial Research (CSIR), Cape Town, South Africa.

Access was granted to the C4 computer clusters at the BioMedical Informatics Centre, Meraka Institute, CSIR, Pretoria, South Africa.

Unnecessary ear damage was prevented by a robust pair of Sennheiser PXC 250 noise reduction headphones.

Many thanks to my family, friends and colleagues in the Malaria Research Group (Department of Biochemistry, University of Pretoria), the Bioinformatics and Computational Biology Unit (University of Pretoria) and the Biochemical Parasitology Group (Bernhard Nocht Institute of Tropical Medicine, Hamburg, Germany). Special mentions go to Ingrid, Jandeli and Marni for your support and friendship.

Summary

Malaria remains a debilitating disease, especially in developing countries of the tropics and sub-tropics. Increasing drug resistance and the rising cost of drug development calls for methods that can cost-effectively identify new drugs. The proteins of the malaria causing *Plasmodium* parasites often exhibit unique features compared to their mammalian counterparts. Such features invite discovery of parasite-specific drugs.

In this study computational methods were applied to understand unique structural features of enzymes from the *Plasmodium* polyamine biosynthesis pathways. Molecular modeling of *P. falciparum* arginase was used to explore the structural metal dependency between enzyme activity and trimer formation. This dependency is not observed in the mammalian host. A novel inter-monomer salt-bridge was discovered between Glu 295 and Arg 404 that helps mediate the structural metal dependency. Removal of the active site metal atoms promoted breaking of the Glu 295_a::Arg 404_b interaction during simulation. The involvement of this salt-bridge was further confirmed by site-directed mutagenesis of the recombinantly expressed enzyme and subsequent simulation of the mutants *in silico*. Mutations designed to break the salt-bridge resulted in decreased enzyme activity and oligomerisation. Furthermore, simulation of the mutants indicated potential loss of metal co-ordination within the active site. The interface around Glu 295_a::Arg 404_b could thus serve as a novel therapeutic target.

In *Plasmodium* the usually separate activities *S*-adenosylmethionine decarboxylase and ornithine decarboxylase occur in a single bifunctional enzyme. Previous studies have established the importance of complex formation and protein-protein interactions for correct enzyme functioning. Disturbing these interactions within the complex may therefore have inhibitory potential. In the second aspect of this study the potential quarternary structure of AdoMetDC/ODC was studied by homology modeling of the domains followed by protein-protein docking. The results from five *Plasmodium* species suggest that one face of each domain is favoured for complex formation. The predicted faces concur with existing experimental results, suggesting the direct involvement of *Plasmodium*-specific inserts in maintaining complex formation. Further fine-grained analysis revealed potentially conserved residue pairs between AdoMetDC/ODC that can be targeted during experimental follow-up.

In both aspects of this study computational methods yielded useful insights into the parasite-specific features of polyamine biosynthesis enzymes from *Plasmodium*. Exploitation of these features may lead to novel parasite-specific drugs. Furthermore, this study highlights

the importance of simulation and computational methods in the current and future practise of Science.

Chapter 1

Introduction

1.1 Polyamines

The polyamines putrescine, spermine and spermidine (Fig. 1.1) are near ubiquitous polycationic aliphatic amines required for a number of essential cellular processes, particularly in organisms undergoing rapid proliferation. Putrescine is typically formed from ornithine by ornithine decarboxylase (ODC) which then serves as a scaffold for the addition of amino-propyl groups from decarboxylated *S*-adenosylmethionine (formed by *S*-adenosylmethionine decarboxylase: AdoMetDC) to produce spermidine (spermidine synthase) and spermine (spermine synthase), respectively (Fig. 1.2). Spermidine and spermine can also be back-converted to their precursors via the combined action of spermine/spermidine N^1-acetyltransferase (SSAT) and polyamine oxidase (PAO). Ornithine is produced from arginine by arginase to release urea. Alternatively, arginine may also serve as the source of putrescine via arginine decarboxylase and agmatine ureohydrolase (Fig. 1.2, Tabor and Tabor, 1984, 1985; Cohen, 1998).

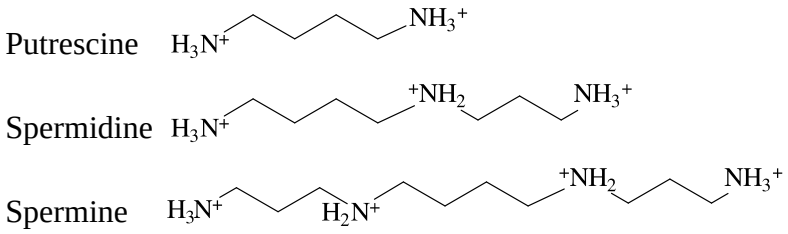


Figure 1.1: The polyamines: putrescine, spermidine and spermine.

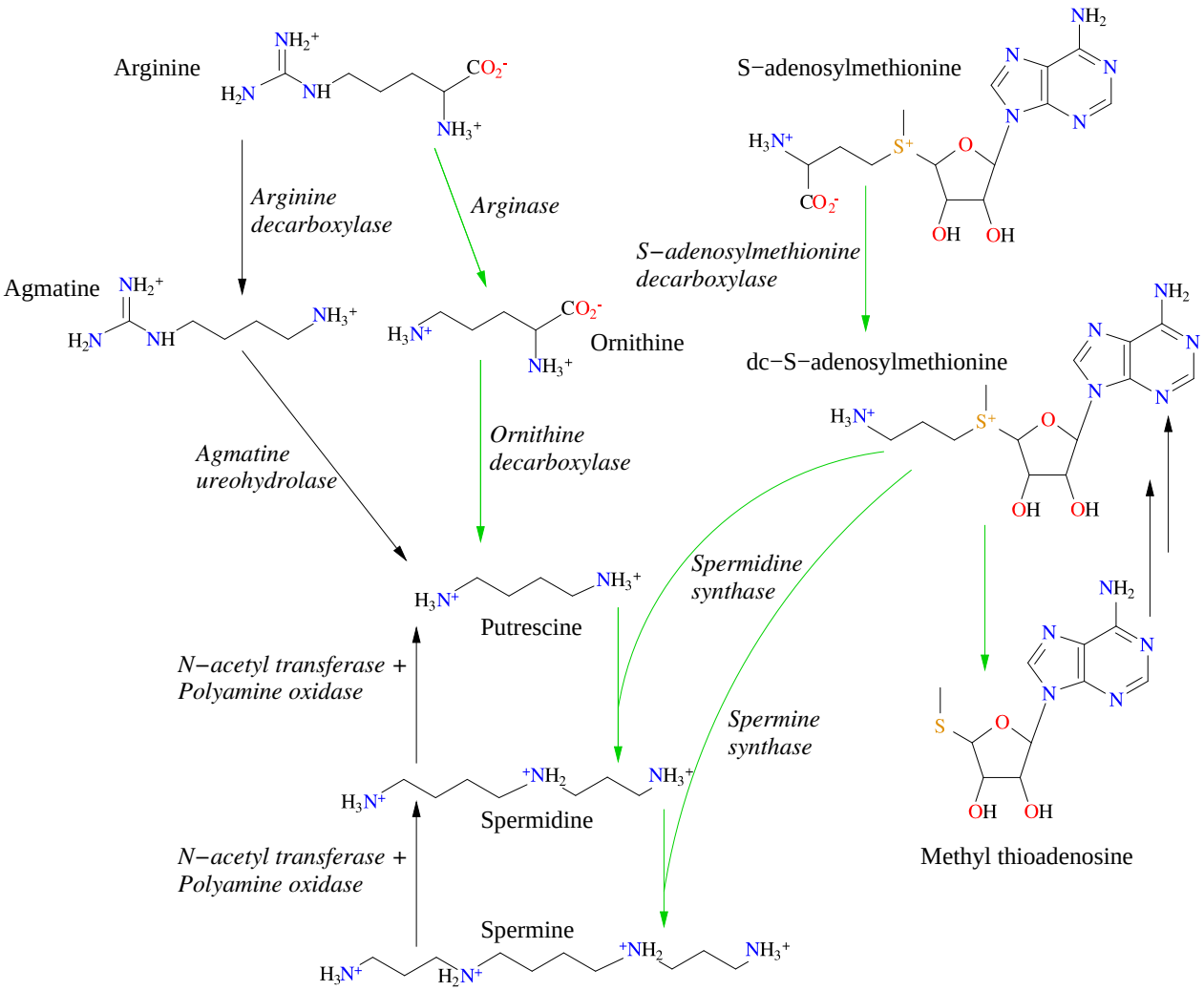


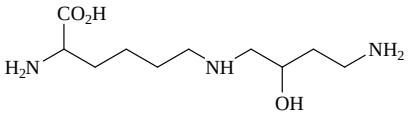
Figure 1.2: Outline of polyamine metabolism. Pathways that have been identified in the malaria parasite *Plasmodium falciparum* are indicated in green. The inclusion of spermine synthase is by virtue of low levels of spermine synthesis by spermidine synthase.

The polycationic nature of polyamines enables them to interact electrostatically with large biological macromolecules such as DNA/RNA and proteins (Bachrach, 2005). It has been suggested that within the nucleus polyamines form aggregates mediated by phosphate ions. These so-called nuclear aggregates of polyamines (NAPs) in turn interact with DNA (D'Agostino and Luccia, 2002; D'Agostino *et al.*, 2005; Luccia *et al.*, 2009). Polyamines can thus affect DNA conformation and chromatin remodeling by enhancing DNA condensation within the tight confines of the nucleus. This in turn affects DNA stability and transcription (Childs *et al.*, 2003; Wallace *et al.*, 2003; Janne *et al.*, 2004).

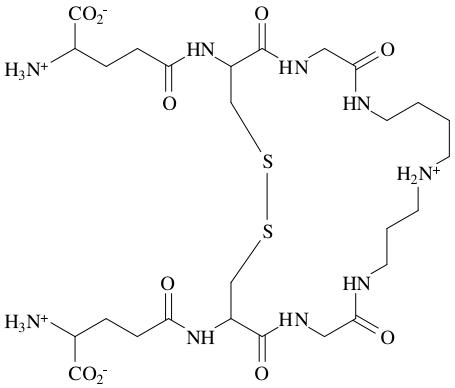
Polyamines are also known to interact with various proteins with varying effects depending on the polyamine species, concentration and protein species. Casein kinase 2 (CK2) interacts with spermine via it's β regulatory subunit leading to enhanced activity. While the biological function of CK2 remains uncertain it has been linked to malignancy via one of its substrates, the oncoprotein Myc. Spermine is also known to modulate the function of membrane proteins such as *N*-methyl-D-aspartate (NMDA) receptors. The formation of protein-DNA complexes is also affected by the presence of polyamines. Complex formation is

typically enhanced, although inhibition may be observed at high polyamine concentrations. Polyamines have also been observed to affect protein degradation, depending on concentration and protein species (Childs *et al.*, 2003).

Additionally polyamines are also required to form certain secondary metabolites, such as the post-translationally modified amino acid hypusine (Tabor and Tabor, 1984; Park *et al.*, 1993) and of the glutathione analogue, trypanothione, in *Trypanosoma* (Fig. 1.3 Müller *et al.* 2003; Heby *et al.* 2007). Hypusine is a post-translationally modified amino acid formed from specific lysine residue of the eukaryotic initiation factor 5A (eIF5A). Firstly, butylamine is transferred from spermidine to the side-chain amino group by deoxyhypusine synthase. This is followed by β-hydroxylation (deoxyhypusine hydroxylase) to form hypusine (Park *et al.*, 1981, 1982). In the protozoan parasites of the trypanosomatid family spermidine is conjugated with glutathione to form trypanothione which replaces the usual glutathione redox system. Glutathione is first conjugated with spermidine to produce glutathionylspermidine by glutathionylspermidine synthetase, followed by a further glutathione conjugation to form trypanothione (Müller *et al.*, 2003; Heby *et al.*, 2007).



(a) Hypusine



(b) Trypanothione

Figure 1.3: Secondary metabolites formed from spermidine

Polyamines also bind to various species of RNA. Binding stabilises tRNA and affects the conformation of 16s rRNA (Amarantos and Kalpaxis, 2000; Amarantos *et al.*, 2002). Furthermore, the presence or absence of polyamines causes translational frame-shifting in certain mRNAs. The Ty1 transposable element in yeast undergoes an increased +1 frame-shift to produce the TYA-TYB fusion protein. During polyamine depletion there is increased +1 frame-shifting and transposition (Clare *et al.*, 1988). In many eukaryotes ODC is regulated by antizyme (AZ) which binds to ODC and targets it for non-ubiquitin mediated proteolysis. AZ mRNA comprises two overlapping open reading frames. In the presence of polyamines the ribosome undergoes a frame-shift to produce the functional protein (Rom and Kahana, 1994).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Tight regulation of polyamines is required for progression through the cell cycle (Ackermann *et al.*, 2003). During the G1 and G2 phases an increase in cellular polyamines is generally observed. Inhibition of polyamine biosynthesis is often observed to arrest cell growth. Furthermore, polyamines are also associated with apoptosis. The association remains uncertain however, with both increased and decreased polyamine levels being linked to both increased and decreased apoptosis. (Wallace *et al.*, 2003).

Polyamine biosynthesis has been identified as a possible therapeutic target for various parasitic diseases (Müller *et al.*, 2003; Heby *et al.*, 2007), cancers (Wallace, 2007) and even HIV via the requirement for hypusine (Schafer *et al.*, 2006). Polyamine biosynthesis enzymes characterised in the malaria parasite *P. falciparum* include the bifunctional *S*-adenosylmethionine decarboxylase/ornithine decarboxylase (Müller *et al.*, 2000; Krause *et al.*, 2000; Wrenger *et al.*, 2001; Birkholtz *et al.*, 2003, 2004), spermidine synthase (Haider *et al.*, 2005) and arginase (Müller *et al.*, 2005).

1.2 Malaria

1.2.1 Introduction and prevalence

Malaria is caused by protozoan parasites of the *Plasmodium* genus and is transmitted by the female *Anopheles* mosquito. Five species of *Plasmodium* are known to infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi* (which till recently was thought to only infect macaques, Singh *et al.* 2004). Of these *P. falciparum* is the most virulent, causing the most deaths. *P. vivax* is the second most dangerous but is only common in tropical regions outside of Africa (Fig. 1.4), the continent with the largest malaria burden (Hyde, 2007; Greenwood *et al.*, 2008). Currently about 2 billion people are at risk of malaria resulting in about 500 million cases annually and 1 million deaths. The majority of the burden exists in developing countries of the tropics and sub-tropics with the majority of casualties being among children. Malaria thus represents an significant impediment to the economic development for much of the world. (Snow *et al.*, 2005; Rowe *et al.*, 2006; Greenwood *et al.*, 2008; Hay *et al.*, 2009).

Plasmodium parasites exhibit a complex life cycle involving a vertebrate host and an invertebrate host (Fig. 1.5). Infection of the vertebrate host begins with inoculation with sporozoites by a mosquito of the *Anopheles* genus. The sporozoites then travel via the blood stream to the liver where they infect hepatocytes. During this asymptomatic stage the sporozoites multiply asexually, eventually releasing themselves from the hepatocyte as merozoites. The merozoites then in turn infect erythrocytes for further rounds of asexual reproduction, passing through various stages (ring → trophozoite → schizont), eventually bursting the red-blood cell to release further merozoites or gametocytes. Free gametocytes are then taken up by another *Anopheles* mosquito, where sexual reproduction occurs (Greenwood *et al.*, 2008).

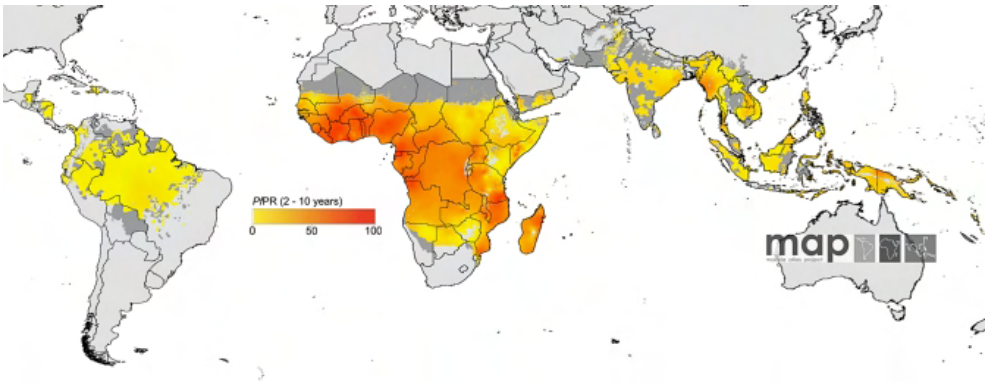


Figure 1.4: Endemicity of *P. falciparum* for 2007 measure as the *P. falciparum* parasite rate (*PfPR*, percentage population with detectable levels of parasites in the blood) for the 2 - 10 year old age group. Adapted from Hay *et al.* (2009).

Global attempts to eradicate malaria beginning in the 1950s achieved partial success in some parts of the world outside of Africa. The two main components of this campaign were the use of chloroquine for treatment and dichloro-diphenyl-trichloroethane (DDT) for vector control. However resistance to both these interventions evolved and the campaign was never attempted in Africa, where there is the highest intensity of malaria transmission. Furthermore, resistance has evolved towards sulphadoxine-pyrimethamine, the front-line treatment that replaced chloroquine (Greenwood *et al.*, 2008).

Current anti-malarials target a number of cellular processes, mostly within the asexual erythrocytic stages of the parasite. Within the cytosol, folate biosynthesis is targeted via inhibition of dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS). Antifolates remains the most common anti-malarial drug class, including pyrimethamine, proguanil, dapsone and sulphadoxine. A DHFR inhibitor is typically used in combination with a DHPS inhibitor, e.g. pyrimethamine and sulphadoxine (fansidar) or proguanil and dapsone. The quinoline family of drugs including quinine, chloroquine, amodiaquine, mefloquine, halofantrine and lumefantrine sequester in the digestive food vacuole of the erythrocytic stages. While their mechanism remains generally unknown it is likely to be mediated by binding haem thus inhibiting haem detoxification. A number of anti-bacterials target the parasite by inhibiting translation within the apicoplast, a chloroplast derived organelle. These include azithromycin, clindamycin and doxycycline. Fosmidomycin acts on the isoprenoid biosynthesis pathway within the apicoplast. Lately artemisinin and it's derivatives have received a lot attention for it's ability as a fast acting drug. Artemisinin is typically combined with longer acting drugs and is notable in that it also targets gametocytes, thus reducing transmission. The mechanism of artemisinin's action remains a topic of considerable debate (Hardman and Limbird, 2001; Greenwood *et al.*, 2008). Resistance has been detected in most drugs in current use (Hyde, 2007). Resistance to artemesinin has been observed in the lab and recent reports indicate this has also emerged in the field on the Thailand-Cambodia border (Dondorp *et al.*, 2009).

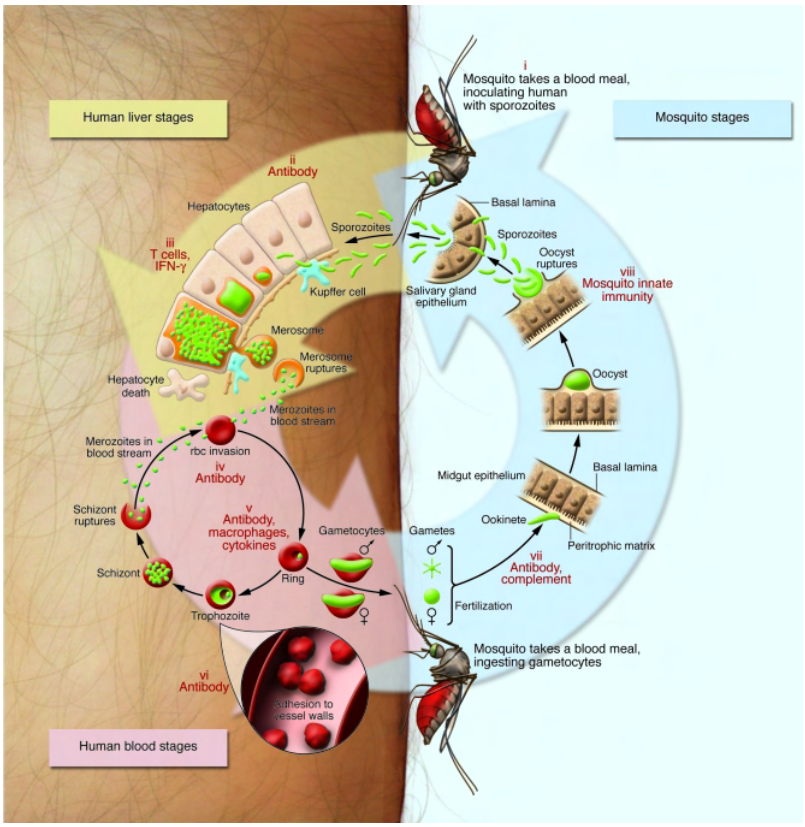


Figure 1.5: Life cycle of *Plasmodium*. Sporozoites are inoculated by a female *Anopheles* mosquito directly into the blood or more often into the dermis (and must then travel to the circulatory system). Upon reaching the liver the sporozoites infect hepatocytes, which rupture and release merozoites approximately one week later. The merozoites then infect erythrocytes to initiate the red-blood cell stages. The parasite then passes through the ring, schizont and trophozoite stages, later rupturing the erythrocyte to release yet more merozoites for further infection. Alternatively, the bloods-stages may develop into the gametocytes to be taken up by an *Anopheles* mosquito during the next blood meal. Adapted from Greenwood *et al.* (2008).

A number of new targets have been identified in recent years, especially as a result of the *Plasmodium* genome sequencing projects. Among these are proteases specific to the digestive food vacuole required for haemoglobin degradation, fatty acid and isoprenoid biosynthesis within the apicoplast, the shikimate pathway as well as lactate, orotate and inositol metabolism (Gardner *et al.*, 2002).

Attempts to generate a malaria vaccine have resulted in mixed success. Vaccination with radiation attenuated sporozoites and sporozoite derived subunits provides partial protection but with waning efficacy. Attempts to immunise against the erythrocytic stages have yet to demonstrate significant protection. In non-human animals (Darwin, 1859) immunisation against the sexual stages has been demonstrated to prevent transmission. While this does not reduce disease it might prove useful in reducing transmission in humans (Greenwood *et al.*, 2008).

Apart from chemotherapy, vector control constitutes the other major arm of malaria eradication. Distribution of insecticide treated nets (ITN) increase survivability in children,

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

while indoor residual spraying (IRS) remains effective in certain areas. However, resistance to pyrethroids, the insecticide most used in ITNs is increasing and DDT only remains effective in restricted areas (Greenwood *et al.*, 2008).

There is thus an urgent need for new anti-malarial strategies if this global problem is to be dealt with.

1.2.2 Polyamine metabolism as a *Plasmodium* drug target

Polyamines are essential for cell growth, proliferation and differentiation. Because of this their metabolism has received a lot of attention as a possible drug target, especially within the cancer research community. Consequently a number of potent inhibitors of polyamine metabolism enzymes have been discovered and developed over the years (Fig. 1.6). Attempts to use these to target polyamines in anti-cancer therapy have largely been disappointing, however. In general, inhibiting polyamine biosynthesis induces cytostasis instead of cytotoxicity within tumour cells (Marton and Pegg, 1995). The reason for the lack of anti-tumour effects is largely due to compensatory mechanisms within the mammalian cell for maintaining the polyamine pools. It has been identified that multiple components of polyamine regulation will have to be targeted when targeting this metabolism for cancer (Seiler, 2003a). Specifically, the uptake of exogenous polyamines by transport systems will likely have to be targeted along with polyamine biosynthesis (Seiler, 2003b). Among the inhibitors identified, the most well known is alpha-difluoromethylornithine (DFMO). DFMO has been used successfully to treat West African Sleeping sickness caused by *T. brucei gambiense* (Wang, 1995). The low toxicity of this drug and its ability to penetrate the blood-brain barrier contribute to it's effectiveness in this regard. The dependence of *Trypanosomes* on the spermidine derived glutathione analogue trypanothione renders them particularly vulnerable to polyamine depletion. The viability of polyamine metabolism as a malarial drug target has been extensively reviewed by Müller *et al.* (2008). DFMO has been found to inhibit *P. berghei* sporogony in *Anopholes stephensi* (Gillet *et al.*, 1983) as well as protect mice from sporozoite infection (Lowa *et al.*, 1986), suggesting that DFMO may have utility as a prophylactic. Targeting the blood stages has proved less promising, however. Inhibition of *Pf*ODC decreases putrescine levels (Wright *et al.*, 1991), while *Pf*AdoMetDC inhibition reduces both spermidine and spermine (Gupta *et al.*, 2005). Cytostasis at the trophozoite stage can be induced by inhibition of *Pf*ODC or *Pf*AdoMetDC but mice infected with *P. berghei* are not cured (Assaraf *et al.*, 1987; Bitonti *et al.*, 1987; Gupta *et al.*, 2005). Co-inhibition of *P. falciparum* ODC/AdoMetDC induces partial transcriptional arrest at the trophozoite stage and transcriptional regulation of proteins in response to polyamine depletion (van Brummelen *et al.*, 2009). Arrest at the trophozoite stage can be rescued by the addition of exogenous putrescine (Assaraf *et al.*, 1987), however, the effect of exogenous spermidine and spermine remains uncertain due to conflicting reports (Assaraf *et al.*, 1987; Wright *et al.*, 1991; Bitonti *et al.*, 1987). Exogenous putrescine is observed to accumulate within the parasites of infected erythrocytes, while erythrocyte levels remain similar to unin-

fected cells. Uptake of exogenous spermidine and spermine has been suggested by the work of Fukumoto and Byus (1996) but also remains uncertain due to conflicting reports (Gupta *et al.*, 2005). Despite these conflicting results, inhibition of polyamine uptake is considered an important target within the parasite. Other polyamine metabolism and related enzymes that have been identified as potential drug targets include *S*-adenosylmethionine synthetase (that produces AdoMet), spermidine synthetase and methylthioadensine (the product of AdoMetDC) recycling enzymes (Müller *et al.*, 2008). To date the most promising result has been obtained when both *Pf*ODC and *Pf*AdoMetDC are inhibited together with exogenous polyamine uptake by the inclusion of bis(benzyl)polyamine analogues. This combined approach was found to be curative of *P. berghei* infected mice (Bitonti *et al.*, 1989). Unpublished results of Brun and Walter with the AdoMetDC inhibitor CGP 40215A produced curative results in *P. berghei* infected mice but without affecting polyamine levels, thus the target remains unclear (Müller *et al.*, 2008). While current results are mixed, polyamine metabolism within *Plasmodium* remains a target worth investigating.

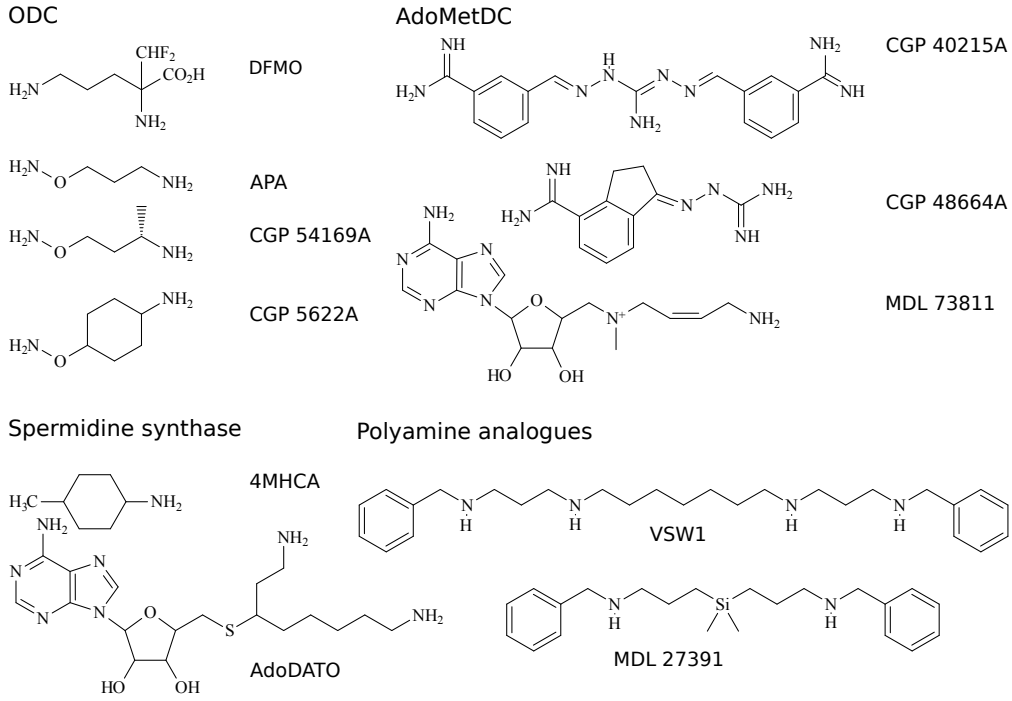


Figure 1.6: Common inhibitors of polyamine metabolism.

1.3 Computational structural biology and rational drug design

1.3.1 Rational drug design

High throughput techniques in modern drug chemistry allow for the generation of large libraries in the order of thousands to millions of compounds. This can be similarly followed up by high throughput screening assays to identify novel lead compounds that may ultimately

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

become or serve as scaffolds for new drugs. Following this approach blindly has only produced modest results, however. Successful high-throughput screening depends on striking a good balance between structural diversity of the compound library while not over-sampling futile regions of chemical space (Lipinski *et al.*, 2001; Snowden and Green, 2008). The average cost of bringing a new drug to market is in the order of $800-900 million (DiMasi *et al.*, 2003; Vernon *et al.*, 2009). Much of the cost results from attrition of lead compounds during the late stages of the research pipeline due to problems with biological availability and toxicity.

A number of computational techniques have been developed in recent years that refine this process and have the potential to substantially reduce the cost of drug discovery. Rational drug discovery depends on having structural information of potential binding compounds and/or the protein target in question. Considering that many drugs act by binding an enzyme active site, the first question that often arises is whether ligand binding can occur for the target protein. A number of docking algorithms exist to predict ligand binding if the structure of both the protein and ligand are at hand. During protein docking various conformational combinations of the ligand and/or protein are sampled and scored in order to determine the most energetically favourable binding. Docking can be applied to whole compounds from large libraries numbering in the millions or to individual fragments which can later be assembled into a larger compound with higher affinity (Klebe, 2006; Orry *et al.*, 2006). Although accurate sampling is computationally expensive, such *in silico* screening is fortunately also embarrassingly parallel and can be distributed across many thousands of computers. While *in silico* docking is not yet absolutely reliable (Ferrara *et al.*, 2004) it can be used to screen out compounds that are unlikely to bind and focus on the most promising candidates.

When the protein structure is absent it remains possible to predict the binding of potential leads by extrapolating from the activities of known ligands and their structures. Using various statistical methods it is possible to extract quantitative structure activity relationships (QSAR) comprising 2D and/or 3D structural descriptors that contribute positively or negatively to activity. These QSARs can then similarly be used to screen the activity of novel compounds (Böhm *et al.*, 1999; Livingstone, 2000). An example of QSARs is given in Figure 1.7.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

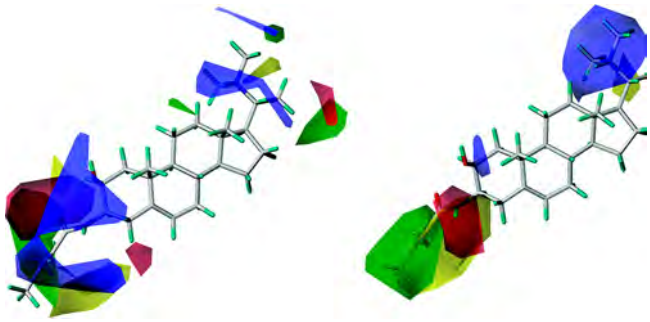


Figure 1.7: Examples of 3D QSARs derived from 39 butyrylcholinesterase inhibitors. A Comparative Molecular field Analysis (CoMFA) QSAR (left) and a Comparative Molecular Similarity Analysis (CoMSIA) QSAR (right). Sterically favourable (green), sterically unfavourable (yellow), positively charged favourable (blue) and positively charged unfavourable (red) regions are depicted. Adapted from Zaheer-ul *et al.* (2008).

In addition to predicting binding it is also important to understand a compound's effect at the organismal level. It is not sufficient for a compound to bind the target in question. A lead should also possess good properties with regards to absorption, distribution, metabolism, excretion and toxicity (ADMET). Unpromising drug leads can therefore be screened out using QSARs with respect to ADMET. Due to the high attrition rate of drug leads at the late stage in development it is becoming essential to screen out as many bad candidates early in the drug discovery process as possible. The most popular of these screens is "Lipinski's rule of 5", whereby it was observed that most sold drugs posses five or less hydrogen bond donors, 10 or less hydrogen bond acceptors, a molecular weight of 500 Da or less and a $c\log P$ of 5 or less (Lipinski *et al.*, 2001).

A good example of the application of rational drug discovery methods for malaria is the so-called World-Wide In Silico Docking of Malaria (WISDOM, http://wisdom.healthgrid.org) project. During the first round of WISDOM about 1 million compounds were docked using FLEXX (Rarey *et al.*, 1996) and AUTODOCK (Goodsell *et al.*, 1996) on two *P. falciparum* plasmepsins using a distributed network on the EGEE (Enabling Grids for E-sciencE, http://www.eu-egee.org) computing grid. A number of known inhibitors were identified as well as a class of novel guanidino based inhibitors from about 40 million dockings. Some if these compounds were later confirmed to be active *in vitro*. During round two dihydrofolate reductase from *P. falciparum* and *P. vivax* as well as glutathione-*S*-transferase from *P. falciparum* were docked against the same ZINC derived library using FLEXX to yield about 140 million dockings. In addition to EGEE a number of other European and affiliated grids were used (Kasam *et al.*, 2009).

1.3.2 Structural modeling

To fully exploit all the methods of rational drug design it is necessary to have the 3D structure of the target protein. The most reliable protein models are generated experimentally via

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

X-ray crystallography and NMR. Current protein structure determination methods are time consuming, however (Bourne and Weissig, 2003). The Protein Data Bank (PDB) currently contains $>$ 57 000 protein structures largely determined using X-ray diffraction and NMR (http://www.rcsb.org). In contrast the number of protein sequences grows much faster. For example just the well-curated SwissProt protein database contains $>$ 500 000 sequences (http://au.expasy.org/sprot/relnotes/relstat.html). Despite the efforts of structural genomics projects which have rapidly increased the pace of protein structure determination, there is a considerable gap between high-throughput structure and sequence data. Newer methods such as high-energy X-ray based methods under development promise to eventually allow for direct determination of protein structure in solution (Mardis *et al.*, 2009; Tiede *et al.*, 2009). Until such techniques become standard computational modeling can fill the gap. Furthermore, for reasons discussed below, *Plasmodium* proteins have proved more difficult than usual to crystallise. For these reasons it is often necessary to follow *in silico* based methods to determine the structures of *Plasmodium* proteins. In this study computational methods are relied on heavily to understand the unique structural features of certain proteins from *Plasmodium* polyamine biosynthesis metabolism.

The holy grail of structural modeling is to be able to predict the 3D structure from sequence alone. The theoretical basis of this is the assumption that the native protein fold is also the global energy minimum of the macromolecule. Therefore, with an accurate mathematical representation of the protein it should be possible to predict the structure by predicting the global minimum. A number of approaches are available for this. The most reliable method would be to model the molecule using computational quantum chemistry. Due to the large computational resources required, however, this approach can only be followed for small molecules (a few hundred atoms) and is generally still not feasible for molecules the size of proteins. Liu *et al.* (2001) have demonstrated with crambin that quantum mechanical simulation is possible using a high performance computing and a semi-empirical quantum representation. It is more common to represent only part of a protein structure quantum-mechanically, e.g. the active site. Although quantum mechanics provides the most accurate description, other methods are required for routine modeling (Leach, 2001; Schlick, 2002).

A more feasible approach is to represent a protein molecule using classical (Newtonian) mechanics. In this treatment the molecule is split up into geometrical components. The terms/components most commonly included are bonds, bond-angles, torsions, improper torsions, electrostatic interactions and Van der Waals interactions (Fig. 1.8). The energy of each component is included in a large sum describing the molecule that can be referred to as the scoring or energy function. The collection of mathematical forms that is used to describe each geometrical component is referred to as the force field. Force fields are typically designed to give a physical description of the molecule in question that could be computed in reasonable time. More complicated force fields include cross-terms to account for interactions between components. Terms can also be derived statistically from distributions of

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

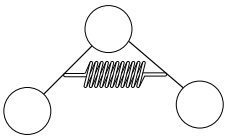
known protein structures and possibly combined with geometric components (Leach, 2001; Schlick, 2002).

$$V = \sum_{bonds} k_b(b - b_0)^2$$

Bonds

- b_0 - reference bond length
- b - bond length
- k_b- force constant

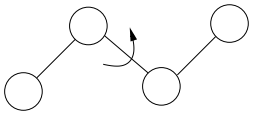
$$V = \sum_{angles} k_\theta(\theta - \theta_0)^2$$



Bond angles

- θ_0 - reference bond angle
- θ - bond angle
- k_θ- force constant

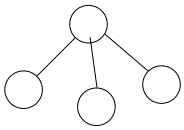
$$V = \sum_{dihedrals} k_\phi(1 - cos(n\phi - \delta))$$



Dihedral/torsion angles

- ϕ - Dihedral angle
- δ - Phase shift
- n - Periodicity
- k_ϕ- Dihedral force constant

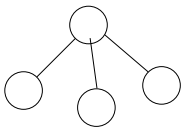
$$V = \sum_{impropers} k_\omega(\omega - \omega_0)^2$$



Improper torsions

- Planer components, e.g. aromatic rings
- $\omega - \omega_0$ - Out of plane angle
- k_ω- Force constant

$$V = \sum_{U-B} k_u(u - u)_0^2$$



Urey-Bradley terms

- Non-bonded atoms connected by 2 bonds:
- Pseudo-bond between 1st and 3rd atoms
- k_ω- Force constant

$$V = \sum_{non-bonded} \epsilon \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}}$$

Non-bonded interactions

- Atoms $>$ 4 bonds apart
- First term: Van der Waals
- Second term: Electrostatic
- r_{ij} - distance between atoms i and j

Figure 1.8: The components of a typical forcefield. Atomistic systems are modelled essentially as "balls on springs" using Newtonian mechanics.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Forcefields are typically used to find an energy minimum (minimisation) as well as for modeling the evolution of a protein in time (molecular dynamics). Using a combination of these approaches structure can be predicted *in silico*. Protein folding typically occurs on a microsecond to second time scale, whereas most molecular dynamics simulations are in the nano-second order. However, with recent advances in methods and hardware it is possible to model proteins *ab initio*. This is more feasible for short sequences but is now becoming tractable for large proteins with adequate hardware resources. Furthermore, simulation runs are being extended into the microsecond time range with millisecond order runs predicted in the near future (Dror *et al.*, 2009; Klepeis *et al.*, 2009).

For the most part, *ab initio* simulations are not possible, however. Instead, starting structures for molecular mechanics-based simulations are generally generated via the method of homology modeling. The basis of this method is to generate a structure via an alignment with a homologue for which the structure is known. Where possible the co-ordinates of the template can be copied for identical portions of the alignment. For non-identical regions semi-empirical methods are used to generate starting co-ordinates. One of the most popular methods is by the satisfaction of spatial restraints, implemented by the MODELLER program. Probability distributions of protein structural features are derived from libraries of existing structures. From these, an objective function can be generated for model structure. Minimisation of the objective function thereby yields a structural model (Fig. 1.9). Closely related to homology modeling is threading (reverse-folding), which is based on the assumption that there are a limited number protein folds. By threading the model sequence through each structure and 3D database and scoring each possibility, remote homology can be detected and in some cases a reliable model structure can be inferred (Bourne and Weissig, 2003; Schlick, 2002).

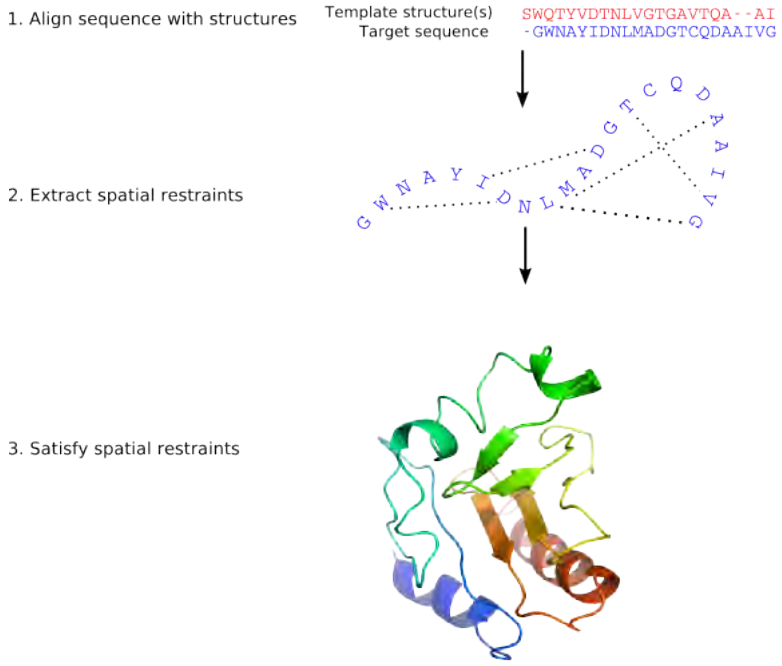


Figure 1.9: Homology modeling is implement in MODELLER.

1.3.3 *In silico* protein-protein docking

Many proteins function through their interaction with one another and other macromolecules such as RNA and DNA. It is predicted that protein-protein interactions are several orders of magnitude larger than the number of protein coding genes. A number of methods exist to predict protein-protein interactions. The yeast two-hybrid system allows for wholesale screening by using potentially interacting proteins to reconstitute a functional transcription factor linked to a reporter gene (Fields and Song, 1989; Stelzl *et al.*, 2005). Techniques designed for testing individual interactions such as affinity purification and chemical cross-linking can be combined with mass-spectrometry for large-scale mapping (Vasilescu and Figeys, 2006). Further experimental methods include tandem affinity purification, synthetic lethality, gene co-expression and protein arrays (Shoemaker and Panchenko, 2007). The number of protein complex structures remains extremely small, however. Wholesale structural determination of protein assemblies require combining a number of techniques (X-ray crystallography, NMR, cryo electron microscopy) and are still low throughput (Lensink *et al.*, 2007). Hence assembly structure determination is likely to lag behind individual structures for some time. The development of computational procedures to predict protein complexes from their individual components are therefore likely to play an important role in filling this gap. A number of non-structural methods exist to predict proteins are functionally and/or physically associated. These include gene neighbour and gene cluster detection, phylogenetic profiling, co-evolution, gene fusion (Rosetta Stone method), classification methods and Bayesian networks (Shoemaker and Panchenko, 2007). In this study the proteins in question are known to interact, the details of these interactions were therefore investigated using

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

structural methods. While computational methods have allowed homology modeling and in some cases *ab initio* structure prediction to become routine, protein-protein docking remains comparatively under-developed. Recent progress has been promising, however, as evidenced by the prediction competition CAPRI (Critical Assessment of PRedicted Interactions). During this competition blind prediction of a protein-protein complex is undertaken by various groups prior to the imminent release of the experimental structure (Méndez *et al.*, 2003, 2005).

A key problem in protein-protein docking is being able to successfully deal with protein flexibility and conformational change upon assembly formation. A successful program must be able to work with individual structures in the so-called unbound conformation to be of any usefulness. Most protein docking algorithms proceed via two stages: rigid docking of the individual proteins followed by further refinement of sidechains and/or protein backbone. Of the most popular methods used for the first stage is the representation of the protein surface on a cubic grid and the use of fast Fourier transforms or geometric hashing to determine geometric complementarity of the protein species (The general procedure is outlined in Fig. 1.10). During this phase all translations and rotations of the so-called mobile species are sampled and scored at predefined distance and angle intervals. Apart from geometric complementarity other features such as electrostatic interactions, Van der Waals interactions, hydrogen bonding and desolvation energies can be included to scoring. During the second stage, high scoring orientations of the first stage can be further refined using methods more akin to classical molecular mechanics based methods. This can include minimisation and molecular dynamics as well as optimisation of side-chain packing from rotamer libraries and backbone remodeling using Monte-Carlo sampling. Protein flexibility can also be handled by the rigid docking of multiple conformations of target protein, e.g. from molecular dynamics (Méndez *et al.*, 2003, 2005; Lensink *et al.*, 2007).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

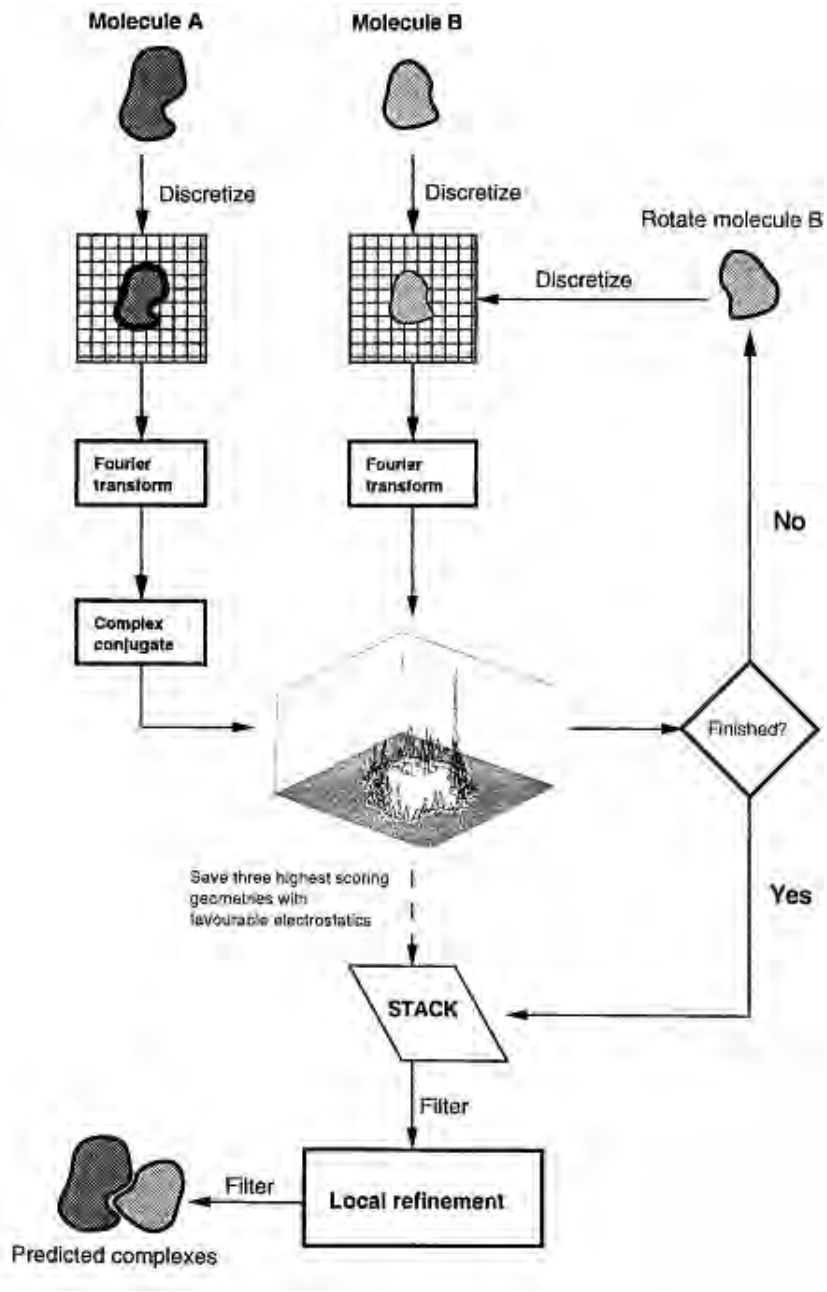


Figure 1.10: A standard protein-protein docking protocol. The molecules are each discretised on a 3D grid followed by the calculation of a correlation function using Fourier transformation. This is used to determine the degree of geometric and electrostatic overlap for a particular orientation. A 3D scan of all possible relative orientations is used to select high scoring positions for further refinement and filtering. Adapted from Gabb *et al.* (1997).

Protein-protein docking is still too unreliable to unambiguously predict the correct relative orientation and exact interface. However, despite the inability to predict the exact orientation, contacting interface residues are often correctly predicted. Furthermore, the various rounds of the CAPRI experiment have demonstrated that incorporation of pre-existing biochemical data often aids in providing accurate predictions. Protein docking predictions can therefore be used to direct course grained experiments such as site-directed mutagenesis

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

of potential interface residues. Predicted complexes can also be used to help solve low resolution structures, e.g. from cryo electron microscopy (Méndez *et al.*, 2003, 2005; Lensink *et al.*, 2007).

Increasingly protein-protein interactions are generating interest as therapeutic targets (Fuller *et al.*, 2009). Targeting protein-protein interactions present certain advantages compared to traditional active site based drugs. Firstly, targeting active sites may not always be the most appropriate approach. Secondly, an active site based drug may have detrimental effects due it's ability to also bind homologous proteins with similar functions. In the case of targeting pathogens, it is desirable to avoid binding to the host cognates for a drug to be effective. While often considered to be undruggable recent successes suggest this may not be the case. The licensed HIV-I drugs enfuvirtide and maraviroc that block viral entry demonstrate this (Melby and Westby, 2009). Proteins from the malaria parasite often possess unique features (discussed below) that make protein-protein interactions an attractive alternative target for novel drugs.

1.4 Malaria proteins as drug targets

1.4.1 Expression of malaria proteins

The rationalised identification of new inhibitors depends on possession of structural information. As for any other organism, the primary problem is obtaining high and pure protein yields for crystallisation trials. Recombinant expression of *Plasmodium* proteins in *E. coli* is notoriously difficult, however. A number of problems are typically encountered. The A+T richness results in substantially different codon usage compared to *E. coli*. *Plasmodium* genes are also typically much longer than their homologues in other organisms, as are the resulting proteins. Increased protein size is due mostly to long protein inserts with generally little homology to cognate enzymes. These inserts tend to be disordered and of low complexity, resulting in proteins that are not amenable to expression and crystallisation. Further problems include sporadic mutations of low complexity sequences introduced by *E. coli*, and cryptic prokaryotic translation start sites within *Plasmodium* genes. Improved levels of protein expression may be obtained by fine control of expression conditions such as a change of strain, addition of rare codon tRNAs or using a completely different expression system. Recently it has become more popular to express the target protein from synthetic genes coding for identical protein sequences but with a codon usage optimised for bacteria (Sugiyama *et al.*, 1996; Withers-Martinez *et al.*, 1999; Yadava and Ockenhouse, 2003; Flick *et al.*, 2004; Christopherson *et al.*, 2004). Mehlin *et al.* (2006) attempted a wholesale expression of 1000 *Plasmodium* genes and obtained soluble expression for only 63 genes. High predicted disorder, molecular weight, *pI* and lack of homology to *E. coli* proteins were all negatively correlated with soluble expression.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

1.4.2 Existing structures

The difficulty of expressing *Plasmodium* proteins is reflected by the paucity of structures in the Protein Data Bank (Kihara and Skolnick, 2003). As of June 2009, querying the PDB (http://www.pdb.org) for structures of *Plasmodium* proteins and excluding sequences with greater than 90% identity, yields 118 entries (de Beer *et al.*, 2009). A closer inspection of all released *Plasmodium* protein structures reveals 100 orthologues from multiple *Plasmodium* species. In contrast, querying the PDB for human protein entries (excluding $>$ 90% sequence identity) reveals more than 4500 structures. Even though the number of *Plasmodium* protein structures is still alarmingly sparse, there has been an almost doubling in *Plasmodium* protein structures since 2005, largely due to the advent of structural genomics programs including the Structural Genomics Consortium, (http://sgc.utoronto.ca) and the Structural Genomics of Pathogenic Protozoa (http:// www.sgpp.org). The Structural Genomics Consortium (SGC) reported 25 distinct *Plasmodium* protein crystal structures from five species. The success rate of this study is similar to other structural genomics programs, and demonstrates the viability of structural genomics for protozoa. This was partly due to treating orthologues from multiple species as alternative expression constructs (Vedadi *et al.*, 2007). The SGPP Consortium has solved 40 structures from the parasitic organisms Leishmania, *Trypanosoma brucei*, *T. cruzi* and *Plasmodium* of which 16 are *Plasmodium* proteins. The success is attributed to pioneering a number of developments such as domain prediction, the use of co-crystallents, capillary crystallisation and "fragment cocktail crystallography".

1.4.3 Modeling of *Plasmodium* proteins

In lieu of the paucity of crystal structures for *Plasmodium* proteins it is often necessary to resort to homology modeling. This approach depends critically on the alignment with template structures. Unfortunately the biased nucleotide and amino acid composition (Bastien *et al.*, 2004b) and *Plasmodium*-specific inserts make it difficult to correctly identify core-conserved regions. The presence of inserts often confuses multiple and structural-alignment programmes. A number of techniques have been used to circumvent this problem (Fig. 1.11). From a first pass alignment, approximate insert positions can be determined. Sequences can then be split according to long inserts and re-aligned. Inserts can vary considerably across different *Plasmodium* species (Birkholtz *et al.* 2004 and C. Claudel-Renard, personal communication). While adjusting an alignment for modeling, it is useful to refer to phylogenetically diverse multiple alignments including as many *Plasmodium* sequences as possible (Wells *et al.*, 2006). As an adjunct to alignment, independent motif discovery (e.g. with MEME, Bailey and Elkan 1994) can be used to fix mistakes that alignment programmes frequently make when aligning long *Plasmodium* proteins with homologues (Wells *et al.*, 2006; de Beer *et al.*, 2006). Further improvements can be made by using hydrophobic cluster analysis (Callebaut *et al.*, 2005) and secondary structure predictions to align homologous regions within inserts. Once an alignment has been decided on, based on visual assessment,

a series of models can be built. Because of the high degree of uncertainty that often accompanies alignments used for modeling *Plasmodium* proteins, it is usually not feasible to rectify all structural anomalies. But by performing standard quality checks on a large sample of models and summarising the results, it is possible to identify parts of the alignment causing most problems. Refined alignments might benefit from species-specific matrices that take into account the differences of amino acid distribution between the aligned proteins (Bastien *et al.*, 2004a, 2005).

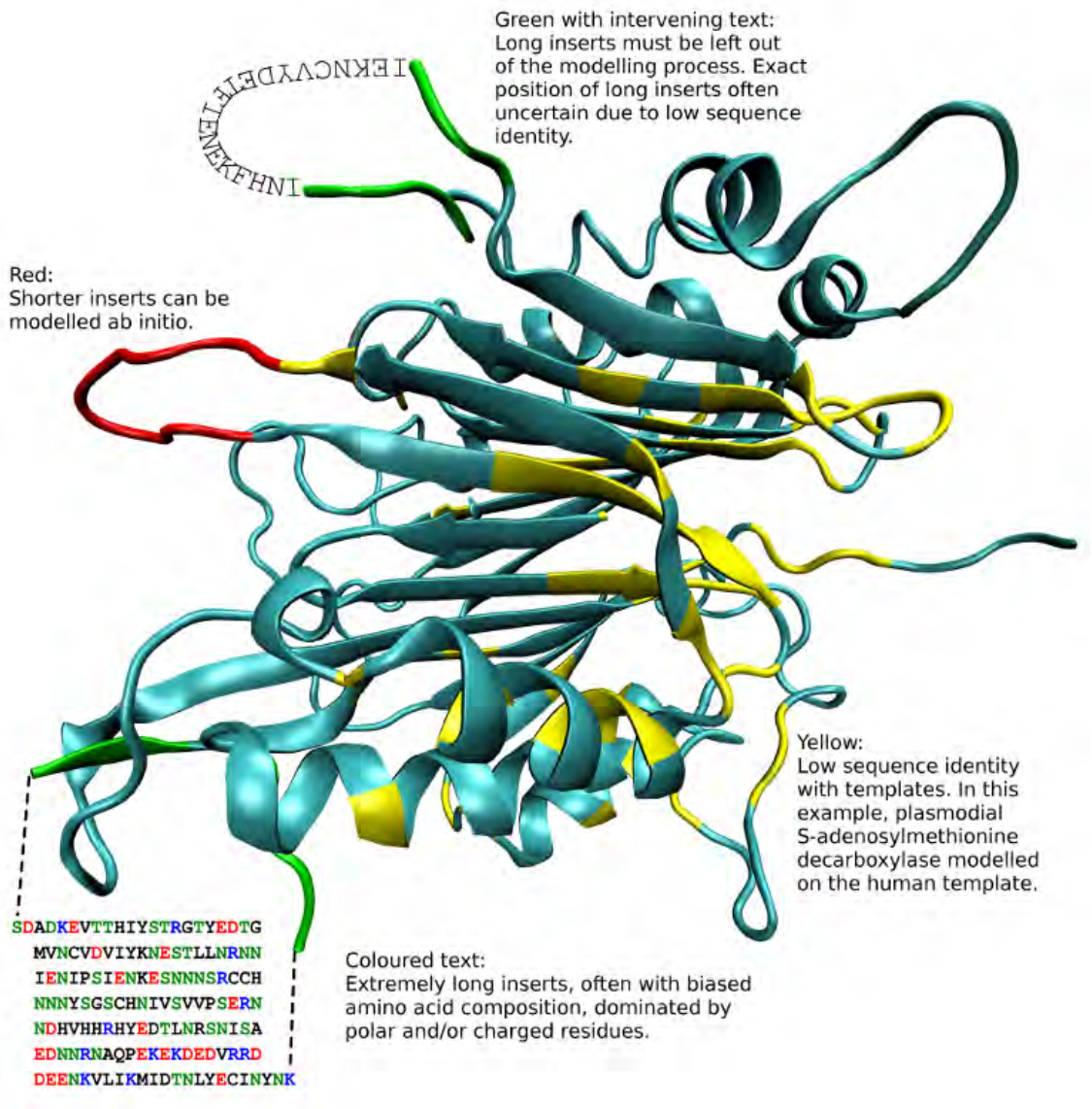


Figure 1.11: Problems frequently encountered with modeling of *Plasmodium* proteins. See text for further details.

Despite the difficulties with homology modeling of *Plasmodium* proteins there have been some notable successes. *P. falciparum* DHFR forms part of a bifunctional protein that also

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

carries thymidylate synthase. A number of existing drugs such as cycloguanil and pyrimethamine target the DHFR domain, and have been used effectively in the past. However, drug resistance has evolved that reduces the usefulness of this important class of drugs. Hence *P. falciparum* DHFR has been a popular target for homology modeling efforts (Toyoda *et al.*, 1997; McKie *et al.*, 1998; Lemcke *et al.*, 1999; Rastelli *et al.*, 2000; Santos-Filho *et al.*, 2001; Delfino *et al.*, 2002). Toyoda *et al.* (1997) were able to identify new inhibitors in the micromolar range. McKie *et al.* (1998) and Lemcke *et al.* (1999) could rationalise the pyrimethamine resistance caused by the Ser 108 Asn mutation. One of these models was further used to identify new inhibitors acting in the nano- and micromolar ranges (McKie *et al.*, 1998). Delfino *et al.* (2002) in turn used their model to investigate a large number of antifolate resistant mutants. Rastelli *et al.* (2000) further explained the cycloguanil resistance/pyrimethamine sensitivity conferred by Ala 16 Val/Ser 108 Thr, as well as the ability for WR99210 to inhibit both pyrimethamine and cycloguanil resistant mutants. A number of new inhibitors were also successfully designed. The high accuracy of the alignment used for modeling meant that predicted dockings were subsequently confirmed with the crystal structure of the complete bifunctional enzyme (Yuvaniyama *et al.*, 2003).

Considerable work has also gone into modeling *Plasmodium* proteases essential to the parasite's intra-erythrocytic life stage. A number of these models have been used to identify new inhibitors (Li *et al.*, 1996; Desai *et al.*, 2004, 2006; de Terán *et al.*, 2006b), although the increasing number of crystal structures for these proteases is likely to gradually replace the need for homology models. de Terán *et al.* (2006a) demonstrated the advantages of using multiple structures with plasmepsin IV from *P. falciparum*. A homology model and a low resolution crystal structure were both used for inhibitor identification. The homology model performed better on structural quality indicators and was more robust when calculating binding energy for an inhibitor series. The enhanced structural quality of the homology model was put down to the intermediate resolution of the X-ray structure (2.8 Å). Further improvements in predicting binding were gained by using a combined model employing both structures, as well as using molecular dynamics to increase sampling. The improved docking performance argues for making use of multiple experimental and predicted models instead of relying on a single structure (Luksch *et al.*, 2008).

Singh *et al.* (2004) used homology modeling to derive a chimeric berghepain-2 that more closely resembled falcipain-2 in it's sensitivity to inhibitors. The motivation behind this approach was to create an *in vivo* rodent model of the *P. berghei* protein that mimics this important human drug target in *P. falciparum*. Homology modeling with molecular dynamics was used to predict the structure, substrate binding and MOA of histo-aspartic protease from *P. falciparum* (Bjelic and Aqvist, 2004). Other noteworthy examples include homology models of dihydropteroate synthase (DHPS) from *P. vivax* and *P. falciparum* to explain the refractory nature of the *P. vivax* enzyme to sulfadoxine (Korsinczky *et al.*, 2004). A homology model of histone deacetylase 1 from *P. falciparum* was successfully used to identify inhibitors in the nanomolar range with significant selectivity compared to mammalian cells (Andrews

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

et al., 2008). Homology models combined with molecular dynamics were used to explain sulfadoxine resistance in mutants of *P. falciparum* DHPS (Rastelli *et al.*, 2000).

A remarkable achievement is exemplified by the homology model obtained for *P. falciparum* farnesyltransferase (Ras FTase) based on a rat homologue (Glenn *et al.*, 2005). The sequence identity between the target and template was quite low (23%) including a parasite-specific insert of approximately 100 residues in the *Plasmodium* protein. Using this model in the docking program GOLD, a range of ethylenediamine based inhibitors with $IC_{50} < 50\ nM$ were identified of which two had an IC_{50} of less than 1 nM. This range of inhibitors was subsequently used together with the model for further rounds of optimisation to derive new structures with better selectivity (up to 145 fold) towards the *P. falciparum* enzyme compared to its mammalian counterpart. Preliminary pharmacokinetics promisingly indicated that some of the compounds were metabolically stable (Glenn *et al.*, 2005, 2006; Fletcher *et al.*, 2008). The results of this work are encouraging and demonstrate that low sequence identity and the presence of inserts need not be a barrier to inhibitor discovery.

1.5 Summary and aims

Due to it's prevalence and increasing drug resistance malaria remains a pressing world health problem that requires urgent attention. Due to increasing resistance, the identification of new drugs remains urgent. This will be best facilitated by a greater understanding of the parasite's basic biology. The high attrition rate of potential drug leads late in the research phase has created the need for rational drug design. This approach will benefit most on gaining structural knowledge of the parasite's macromolecules that represent promising targets for inhibition. However, *Plasmodium* proteins possess a number of characteristics that make structural determination difficult using current experimental methods. To fill the gap computational methods can be applied to facilitate further experiments designed to understand and possibly exploit these unique characteristics. The enzymes of the *P. falciparum* polyamine pathway have been identified as a potential drug target and also exemplify this unique characteristic of *Plasmodium* proteins. Compared to the human host, the arginase of *P. falciparum* displays a strong dependency between trimerisation, enzyme activity and the presence of the active site metals. Additionally, the bifunctional arrangement of *S*-adenosylmethionine decarboxylase/ornithine decarboxylase is apparently unique to *Plasmodium* and is definitely absent in the human host. This study describes the application of various molecular modeling techniques to further understand the unique characteristics of these enzymes. Homology modeling and molecular dynamics of arginase revealed a novel inter-monomer interaction that is involved in the structural metal dependency. This interaction may serve as a potential parasite-specific target. Homology modeling and docking of AdoMetDC and ODC from five *Plasmodium* species was pursued to predict the quaternary structure of the bifunctional complex. Conserved regions and specific residues were identified as likely candidates for mediating AdoMetDC/ODC binding, and have targeted for further experimental follow-up.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The findings discussed can and have been used to guide further experimental analysis that may ultimately lead to novel therapeutic exploitation of these proteins.

Partof this work has been published in the FEBS Journal (Wells *et al.*, 2009) and presented at the following conferences:

- Investigations into the structural metal dependency of malarial arginase with molecular dynamics. Intelligent Systems for Molecular Biology (ISMB). July 2007, Vienna, Austria. 2.

- Investigations into the structural metal dependency of malarial arginase with molecular dynamics. First African Structural Biology Conference. November 2006, Wilderness, South Africa.

Chapter 2

Investigation of the Structural Metal Dependency of *P. falciparum* Arginase with Molecular Dynamics

2.1 Introduction

2.1.1 Structure and reaction mechanism

Arginase (L-arginine-urea hydrolase, EC 3.5.3.1) catalyses the hydrolysis of (L-)arginine to (L-)ornithine and urea (Fig. 2.1). The arginase fold also includes agmatinase (Ouzounis and Kyrpides, 1994; Ahn *et al.*, 2004) which converts agmatine to putrescine and urea. Agmatine is formed by decarboxylation of arginine (arginine decarboxylase). Arginase is thus part of one of two alternative biosynthetic routes to putrescine (Fig. 1.2). Arginase and agmatinase form part of the ureohydrolase superfamily, that also includes histone de-acetylase and acetyl-polyamine amidohydrolase (Dowling *et al.*, 2008). This superfamily is considered to be a very ancient fold that was probably present in the last universal common ancestor (Ouzounis *et al.*, 2006). Arginase is thought to have evolved later from the agmatinase fold since the presence of ornithine early in evolution may have been detrimental due to mis-incorporation into proteins (Sekowska *et al.*, 2000).

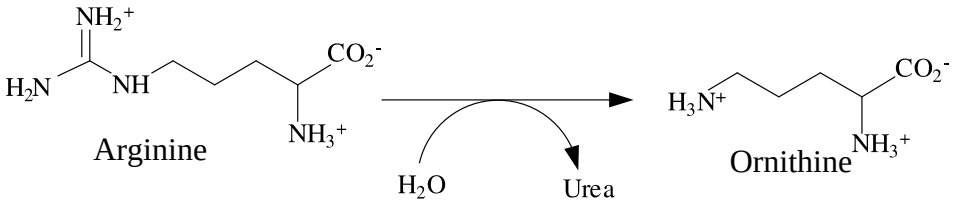


Figure 2.1: Arginase reaction

Arginase is a multimeric metallo-enzyme with a binuclear spin-coupled Mn^{2+} cluster in the active site. A number of structures have been determined for arginase. Among these are structures from the bacteria *Bacillus caldovelox* (Bewley *et al.*, 1996, 1999), human arginase I (Di Costanzo *et al.*, 2005) and II (Cama *et al.*, 2003a), and rat arginase I (Kanyo *et al.*, 1996; Scolnick *et al.*, 1997). A structure has also been deposited in the PDB for arginase from *Thermus thermophilus* (PDB ids: 2EF4, 2EF5, 2EIV). Each active site is restricted to a single α/β monomer. The monomer comprises an eight stranded parallel β-sheet packed between α-helices on either face. The β-strands are arranged in the order $\beta2 - \beta1 - \beta3 - \beta8 - \beta7 - \beta4 - \beta5 - \beta6$ (Fig. 2.2).

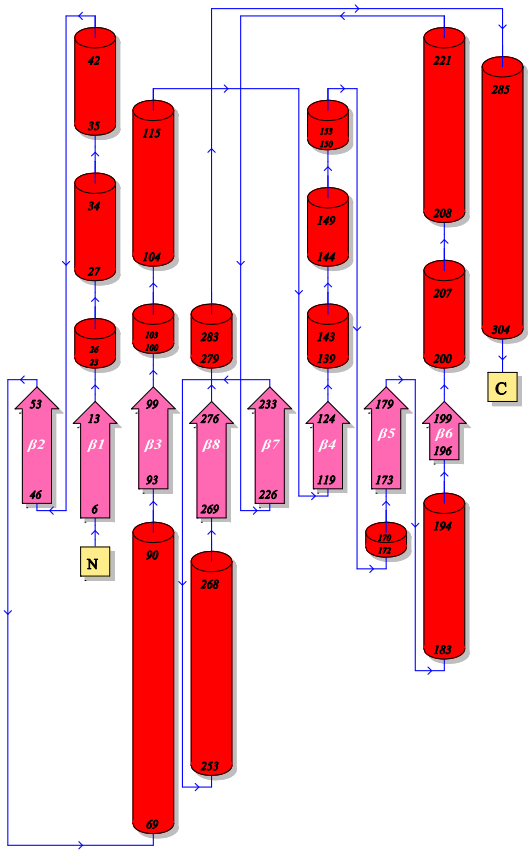


Figure 2.2: Topology of the monomer from rat arginase I (PDB id: 1RLA). Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The figure was adapted from output generated on the PDBsum server (http://www.ebi.ac.uk/pdbsum/).

The metal cluster sits in the active site in a 15 Å deep cleft at the *C*-terminal end of the β-sheet. The Mn^{2+} atoms are about 3.3 Å apart and are bridged by a solvent molecule (Kanyo *et al.*, 1996; Bewley *et al.*, 1999). Mn_A^{2+} is the more inaccessible of the two ions and was initially reported to be co-ordinated by five ligands in rat arginase I with a distorted square pyramidal geometry. The co-ordinating ligands comprise one His residue, three monodentate Asp residues and the bridging solvent. The second ion (Mn_B^{2+}) was reported to be co-ordinated by six ligands in a distorted octahedral geometry (Scolnick *et al.*, 1997), including one His, two monodentate Asp residues, one bidentate Asp residue and the bridging solvent. The two Mn^{2+} ions are bridged by three ligands: one bidentate Asp, one monodentate Asp and the bridging solvent. X-ray absorption spectroscopy suggested both ions are co-ordinated octahedrally (Stemmler *et al.*, 1997), however, with an extra solvent molecule binding to Mn_A^{2+}. This was confirmed by the bacterial structure (Bewley *et al.*, 1999) and re-examination of the original rat I arginase structure (Kanyo *et al.*, 1996) (Fig.

2.3). The correspondence of the rat I, human II and bacterial Mn^{2+}co-ordinating residues is given in Table 2.1. The requirement of these ligands for enzyme activity have been confirmed by a number of mutagenesis studies. Mutation of His 101 and His 126 to Asn in rat arginase I compromise activity (Cavalli *et al.*, 1994). The first of these disrupts Mn_A^{2+} and thermostability (Scolnick *et al.*, 1997). Furthermore, the mutations Asp 233 Ala/Cys, Asp 128 Glu/Asn/Ala, Asp 234 Glu/Ala/His and His 101 Glu all compromise activity and Mn^{2+} binding (Cama *et al.*, 2003b).

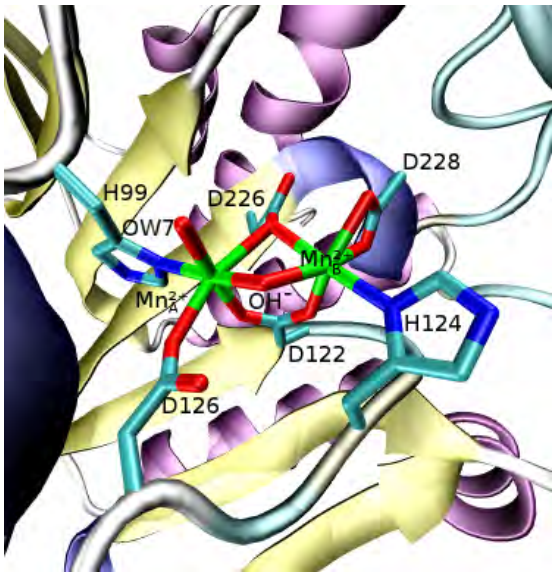


Figure 2.3: Co-ordination pattern of Mn^{2+} in bacterial arginase (PDB id: 2CEV). Both Mn^{2+} ions display octahedral co-ordination, however, in some cases Mn_A^{2+} shows square-pyrimadal co-ordination (The water molecule OW7 is absent).

Table 2.1: Correspondence of Mn^{2+} co-ordinating and other active site residues in rat I-, human II- and bacterial- (*Bacillus caldevelox*) arginase.

Rat I	Human II	Bacterial	Rat I	Human II	Bacterial
His 100	His 120	His 99	His 141	His 160	His 139
Asp 124	Asp 143	Asp 122	Asp 232	Asp 225	Asp 226
His 126	His 145	His 124	Asp 234	Asp 227	Asp 228
Asp 128	Asp 147	Asp 126			

In the mechanism proposed by Di Costanzo *et al.* (2005) the metal bridging solvent is an activated hydroxyl which attacks the guanidinium carbon of arginine, followed by collapse of the tetrahedral intermediate to release ornithine and urea (Fig. 2.4). However, *ab initio* modeling of the active site suggests that the bridging solvent may be a neutral water molecule instead (Ivanov and Klein, 2004). The high resolution structure of human arginase I reveals that His 141 is protonated and most probably acts as a general acid to

protonate the leaving group of L-ornithine. The catalytic role of this residue is confirmed by mutagenesis to Asn (Cavalli *et al.*, 1994) and Phe (Carvajal *et al.*, 1999a) which substantially reduce activity but have no effect on Mn^{2+}binding or K_m. The mechanism is confirmed by the binding of various inhibitors that are expected to mimic the tetrahedral transition state. Among these are the borate ion $B(OH)_4^-$(Carvajal *et al.*, 1999b), inhibitors with the trigonal borono ($-B(OH)_2$) moiety which is attacked by the nucleophillic OH^- to form a tetrahedral intermediate ($-B(OH)_3^-$) and sulphonamides (Cox *et al.*, 2001; Cama *et al.*, 2003a,c, 2004). The metal-bridging by the carbonyl moiety of urea is confirmed by a similar binding mode for the C = S group of thiosemicarbazide (Di Costanzo *et al.*, 2005).

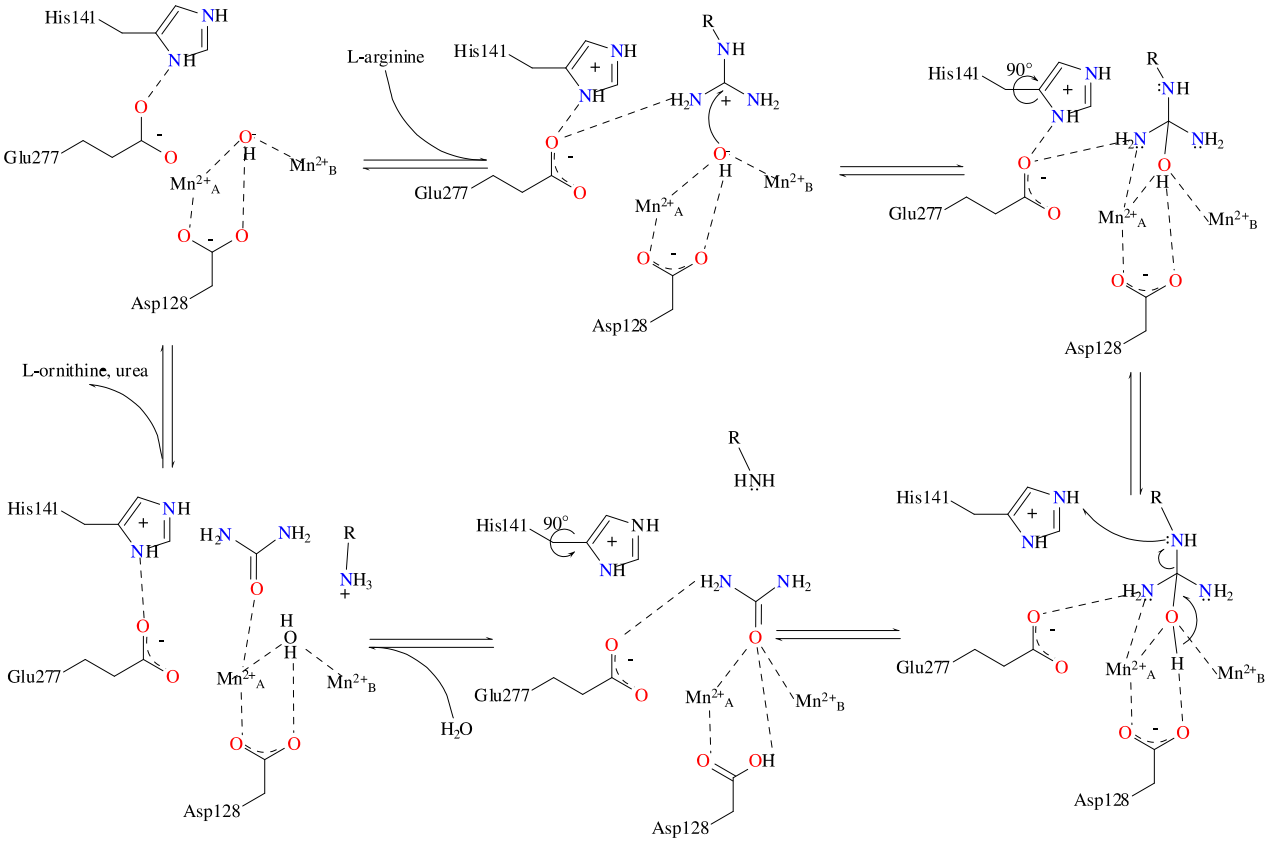


Figure 2.4: General reaction mechanism of arginase, adapted from Di Costanzo *et al.* (2005). Residue numbers correspond to rat I arginase. The activated hydroxyl attacks the guanidinium carbon of arginine, followed by collapse of the tetrahedral intermediate to release ornithine and urea. See main text for further details.

2.1.2 Quaternary structure

In most recent literature arginase is observed within trimeric or hexameric complexes. Carvajal *et al.* (1971) did report human arginase I as being tetrameric, however, this is probably because the molecular mass of arginase was underestimated to be 30 kDa, as opposed to the true molecular mass of 34.7 kDa (Di Costanzo *et al.*, 2007a). Eukaryotic arginases are observed to be trimeric, whereas bacterial arginases are hexameric (Jenkinson *et al.*, 1996). Furthermore, in *S. cerevisiae* trimeric arginase forms a hetero-hexameric complex

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

with trimeric ornithine transcarbamoylase. Ornithine transcarbamoylase catalyses the first committed step in arginine biosynthesis, and occurs in the cytosol together with arginase in yeast. The presence of both these enzymes could potentially keep urea in a futile cycle leading to hydrolysis of ATP, however, when arginase complexes with and inactivates ornithine transcarbamoylase (Bechet and Wiame, 1965; Messenguy and Wiame, 1969).

In human arginase I the main structural determinant for oligomerisation was found to be an inter-monomer salt-bridge between Glu 256 and Arg 255. Mutating Glu 246 to Gln abolished trimer formation, with no effect on kinetic properties (Sabio *et al.*, 2001). This salt-bridge has been observed in all arginase structures so far, and is likely to be necessary in all arginases for oligomerisation.

In the first structure of rat arginase I Kanyo *et al.* (1996) reported that mostly inter-monomer contact was made by a novel S-shaped motif at the *C*-terminus. This was tested in human arginase I by deletion of fourteen residues following Arg 308 (= rat I Arg 308) which apparently nucleates the interactions of the S-shaped motif by forming a salt bridge with Asp 204. The *C*-terminus was not found to be required for oligomerisation of human arginase I. Furthermore, treatment with high salt concentrations did not disrupt the oligomer, suggesting that Arg 308 is also not required for oligomerisation. The kinetic properties were also unaltered (Mora *et al.*, 2000). The role of Arg 308 was later contradicted somewhat by Lavulo *et al.* (2001), where mutating Arg 308 to Ala, Glu or Lys in rat arginase I abolished trimer formation. These mutations were also found to compromise enzyme activity where k_{cat} was 33% - 41% of the wild type. The S-shaped tail is conserved in multiple isoforms of eukaryotic arginase (Lavulo *et al.*, 2001). In bacterial arginase from *Bacillus caldevelox* Arg 308 aligns with Leu instead. Furthermore, at the inter-monomer interface a guanidinium ion is observed at the corresponding position for rat I Arg 308 when crystals are grown in guanidinium hydrochloride (Fig. 2.5). This guanidinium may be replaced by free arginine, depending on the crystallisation conditions (Bewley *et al.*, 1999).

Previously, a relationship between active site metal binding and tertiary and quaternary structure in arginase has been reported. Carvajal *et al.* (1971) have reported that human arginase I requires Mn^{2+} in order to form oligomers. However, active monomers immobilised on nylon could be formed by incubation with Mn^{2+} (Carvajal *et al.*, 1977, 1978, 1982). Arginase from *Saccharomyces cervisiae* was initially reported to contain a weakly bound $Mn^{2+}$and a strongly bound Zn^{2+} ion. Removal of the strongly bound ion caused dissociation of the trimer and loss of activity. *S. cerevisiae* arginase requires a heating step at 45 °C in the presence of $Mn^{2+}$to become active. Removal of both the weakly and strongly bound ions caused a decrease in T_m from 57 °C to 54 °C and 38 °C, respectively (Green *et al.*, 1990, 1991).

Thus, in mammals substantially active monomers are still formed by mutations that abolish trimer formation (Lavulo *et al.*, 2001; Sabio *et al.*, 2001). Furthermore, mutations disrupting metal binding has no reported effect on the quaternary structure (Cavalli *et al.*, 1994; Scolnick *et al.*, 1997; Cama *et al.*, 2003b). In contrast, a stronger relationship between

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

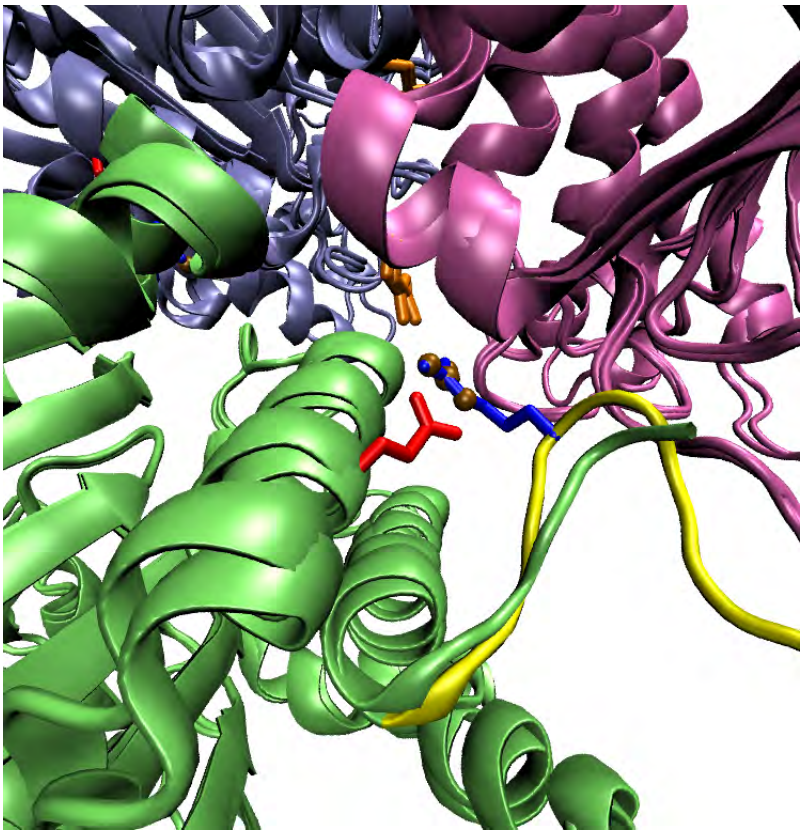


Figure 2.5: Inter-monomer interactions in rat arginase I (PDB id: 1RLA) and *B. caldovelox* arginase. Monomers are in green, light-blue and mauve. The S-shaped tail of rat arginase I is indicated in yellow. Asp204/199 (rat/*B. caldovelox*) is indicated in orange, and Glu256 of *B. caldovelox* in red. Arg308 of rat arginase I is indicated in blue, and guanidinum from *B. caldovelox* in brown. The Arg308 interaction from rat arginase I is replicated by the guanidinium bridging between Asp199 and Glu256 in *B. caldovelox*

metal binding and quarternary structure has been observed in *P. falciparum* (Müller *et al.*, 2005). Mutation of *Pf* Arg Glu 347 to Gln causes dissociation of the trimer as well as inactivation of the enzyme by 90%. This contrasts with human arginase I, where Glu 256 Gln is fully active. Furthermore, mutations to the active site that compromise activity also cause the trimer to dissociate (Table 2.2). This has yet to be reported in other species.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 2.2: Properties of mutated *P. falciparum* arginase in comparison to the wild type enzyme.

Name	Activity (% of wild type)	Structure
Wild type	100	Trimer
His 193 Ala	Inactive	Trimer
His 218 Ala	Inactive	Monomer
His 233 Ala	Inactive	Monomer
Asp 216 Ala	Inactive	Monomer
Asp 220 Ala	Inactive	Monomer
Asp 323/325 Ala	Inactive	Monomer
Glu 347 Gln	10	Monomer

2.1.3 Metabolic functions

In mammals, two isoforms of arginase have been identified. Arginase I is cytosolic and largely hepatic where it catalyses the final step of the urea cycle (Spector *et al.*, 1982, 1985). Non-hepatic arginase I is likely to be involved in the regulation of NO synthesis in immune tissue. Inducible NO synthase (iNOS) and arginase both compete for arginine as a substrate. These are reciprocally regulated by Th1/Th2 (Munder *et al.*, 1999), and arginase is inhibited by NOHA (N^{G}-hydroxy-L-arginine), an intermediate of iNOS (Hecker *et al.*, 1995). Arginase II is non-hepatic, and occurs in the mitochondrial matrix (Grody *et al.*, 1987; Gotoh *et al.*, 1996; Cox *et al.*, 2001). Arginase II is involved in homoeostasis of ornithine for production of proline and glutamate (Morris, 2002) and is also potentially involved in regulating NO biosynthesis (Li *et al.*, 2001; Cama *et al.*, 2003a). Bacterial arginases have a single isoform, while more than one exists in yeast (Borkovich and Weiss, 1987). In the yeast *Nuerospora crassa* two isoforms have been reported from a single gene under control of tandem promoters. Yeast arginase has been implicated in glutamate accumulation during germination and asexual spore development (Marathe *et al.*, 1998; Turner and Weiss, 2006). Arginase in *P. falciparum* is involved in the biosynthesis of the polyamine putrescine and is likely to be the sole biosynthetic route to ornithine, since no agmatinase has been identified in the parasite (Müller *et al.*, 2005).

2.2 Aims

P. falciparum arginase exhibits an interesting metal/trimer dependency compared to the human host. Mutations that abolish metal binding or removal of the metal ions by dialysis and EDTA chelation causes dissociation of the trimer into inactive monomers. Conversely, mutations that abolish inter-monomer interactions which are far from the active site result in inactive monomers (Müller *et al.*, 2005). This host-parasite difference may thus provide a novel non-active site based strategy for inhibiting *P. falciparum* arginase. Thus the mechanism of this structural dependency was investigated by homology modeling and molecular

dynamics, in order to establish an *in silico* system for exploiting this dependency.

Specifically, model systems of *P. falciparum* arginase were simulated with and without the active site metal as well as various mutants predicted to affect trimer interaction. The effects of these mutations were also tested experimentally on recombinantly expressed protein.

2.3 Methods

2.3.1 Sequence alignments

Reference multiple alignments were generated using CLUSTALW 1.82 and the FUGUE (Shi *et al.*, 2001) server. Default parameters were used in both cases. The CLUSTALW alignment included Eukaryotic Arginase types I and II and bacterial Arginases. The following sequences were used for the CLUSTALW alignment (Entrez accession number are given in brackets for non-*Plasmodium* species, PlasmoDB reference numbers are used for *Plasmodium* sequences): *Arabidopsis thaliana* (P46637, Q9ZPF5), *Schizosaccharomyces pombe* (P37818, Q10066), *Xenopus laevis* (Q91553, Q91554, Q91555, P30759), *Homo sapiens* (P78540, P05089), *Mus musculus* (O08691, Q61176), *Rattus norvegicus* (O08701, P07824), *Agrobacterium tumefaciens* (P14012), *Bacillus caldovelox* (P53608), *Bacillus subtilis* (P39138), *Brucella melitensis* (Q59174), *Coccidioides immitis* (P40906), *Emericella nidulans* (Q12611), *Neurospora crassa* (P33280), *Rana catesbeiana* (P49900), *Glycine max* (O49046), *Staphylococcus aureus* (P60086), *Saccharomyces cerevisiae* (P00812), *Plasmodium knowlesi* (PKH_070380), *Plasmodium vivax* (Pv098770), *Plasmodium falciparum* (PFI0320w), *Plasmodium yoelii* (PY03443), *Plasmodium berghei* (PB000787.03.0). Sequences for *P. knowlesi*, *P. vivax*, *P. yoelii* and *P. berghei* were obtained by using BLAST as provided on the plasmodb.org website with the *P. falciparum* sequence as query. Although the reference alignments were often highly redundant, all sequences were retained to offset the bias of including five *Plasmodium* sequences.

2.3.2 Homology modeling

MODELLER 7v7 and 8v0 were used to build homology models. Trimeric models were constructed on the rat arginase I (PDB id: 1RLA[abc], Kanyo *et al.* (1996)), human arginase II (PDB id: 1PQ3[abc], Cama *et al.* (2003a)) and *Bacillus caldovelox* (PDB id: 1CEV[abc], Bewley *et al.* (1999)) templates. Super-imposable monomers were constructed by imposing symmetry restraints on the internal co-ordinates of all atoms during the model building process (MODELLER 8 only). Models were built using very slow refinement. The effect of various sequence alignments was determined by generating multiple models with different random number seeds and monitoring the effect on the number of residues in disallowed regions of the Ramachandran plot and on the overall G-factor score from PROCHECK. Problem areas were identified as residues that frequently fell in disallowed regions. Models that minimised residues with poor phi/psi values and maximised the G-factor were used for molecular dynamics.

Three alignments were used to generate homology models. The first of these (a1, Fig. 2.6) was initiated with a CLUSTALW alignment, followed by manual adjustment with reference to existing crystal structures. The second alignment was obtained using the FUGUE structural alignment server, followed by manual adjustments (a2, Fig. 2.7). Two parasite-specific inserts occur in *Pf* Arg. The third alignment (a3) was created by moving the position of insert 2 in a1 to that of a2 (one residue downstream). In all cases the full trimer was modelled. In a2 and a3 the first 16 residues were excluded to avoid the uncertainties involved with modeling long stretches *ab initio*.

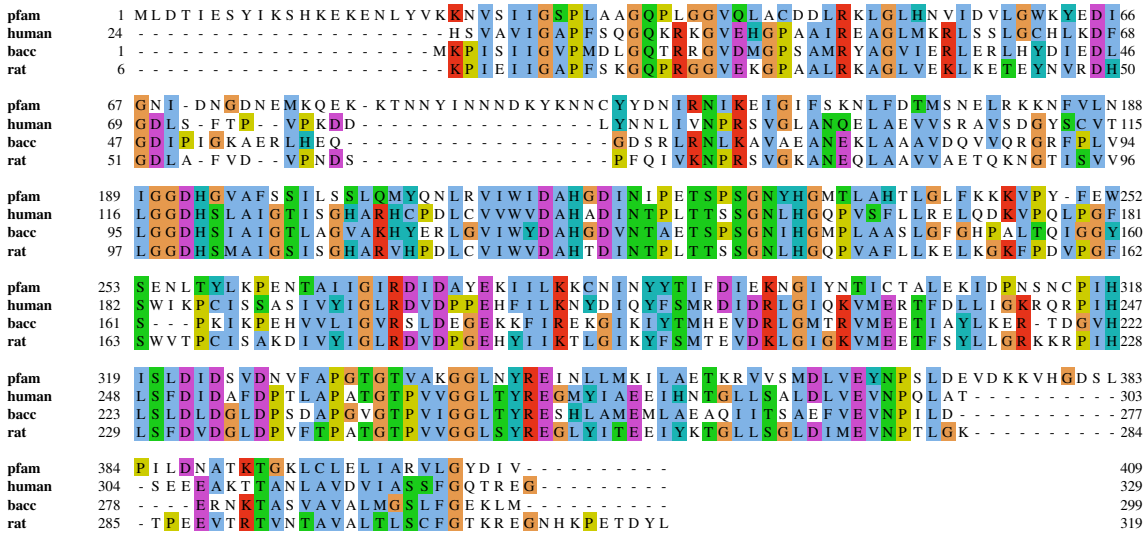


Figure 2.6: Alignment 1 (a1). Constructed with CLUSTALW multiple alignment as reference. Amino acid similarity is indicated using the CLUSTAL (Thompson *et al.*, 1997) colour scheme.

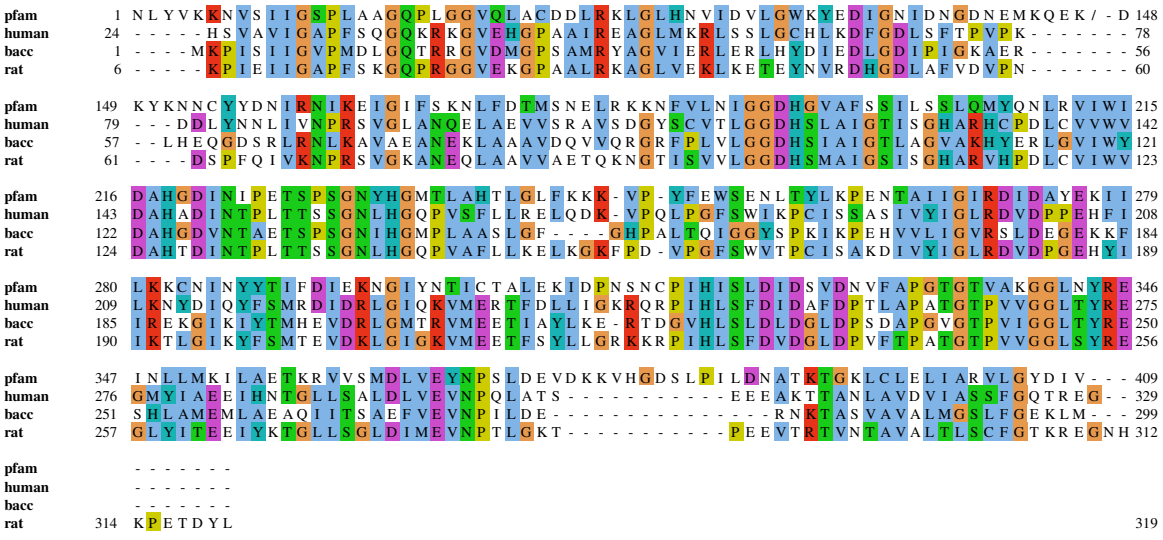


Figure 2.7: Alignment 2 (a2). Constructed with FUGUE alignment as reference. In alignment 3, insert 2 of this alignment is as in alignment 1. Amino acid similarity is indicated using the CLUSTAL (Thompson *et al.*, 1997) colour scheme.

2.3.3 Molecular dynamics

The methodology followed was essentially an exploratory approach as well as a learning process. Different protocols and software were employed in order to find the most appropriate *in silico* system for the hardware available, at the same time producing results that agree with previous experiments, especially with regards to known protein-protein interactions.

2.3.3.1 Simulations with CHARMM

System setup: Hydrogen atoms were added automatically using CHARMM (29B2 or 32B1). The histidine protonation scheme adopted is based on known requirements for catalytic activity and co-ordination of metal atoms in the active site. Thus His 193 and His 218 were protonated on N_ϵ. All other histidines were protonated on N_δ. Glu, Asp and Lys residues were charged. Since the current CHARMM forcefields are not parameterised for Mn^{2+}, the Mg^{2+} ion was used instead. It was thus assumed that any effects of the metal on trimer formation were largely electrostatic in nature (discussed below).

Long range interactions: For CHARMM simulations the same scheme was used to calculate non-bonded interactions (van der Waals and electrostatic) during minimisation and molecular dynamics. Typically, interactions were calculated between all atom pairs, except atoms connected by three or less covalent bonds (1-2, 1-3, 1-4 interactions). A constant dielectric was used, to simulate solvent attenuation of both non-bonded interactions. Non-bonded interactions were shifted to zero at from 8 Å to a cut-off of 12 Å (beyond which long-range interactions are ignored) to prevent the introduction of numerical errors.

Heating and equilibration:

Solvated shell system: An early homology model (methods not shown) was used to create a shell solvated trimer. The protein trimer was first minimised for 20 steps with steepest descent (SD). The positions of the atoms were restrained with harmonic force constraints of 20 ($\times 2 \times$MW $kcal.mol.\AA^2$). This was followed by minimisation with conjugate gradient (CJ) in various stages with a decrease in the force constant on the harmonic restraints (10: 200 steps, 5: 200 steps, 0: 100 steps). The system was then explicitly solvated with the TIP3 water model in a shell of 10 Å (55602 atoms). The whole system was then minimised with a harmonic force constant of 50 on the solute (protein) for 50 steps with steepest descent and 100 steps with conjugate gradient. The solvent was then heated (force constant 50 on solute) from 60 K to 298 K in 10 K increments every 500 steps using velocity scaling. The SHAKE algorithm was used to restrain hydrogen atom movement. The timestep was 1 fs and the run lasted 30 000 steps (30 ps).

***In vacuo* trimer - early model:** The same model as used for the solvated shell system was used for molecular dynamics *in vacuo*. The system was then minimised in various stages

as previously described. Harmonic restraints were applied to all atoms. The stages were as follows: 20 steps SD - force constant 20, 200 steps CJ - force constants 10, 200 steps CJ - force constants 5, 100 steps CJ - force constants 0. The system was then heated from 58 K to 298 K in 10 K increments every 500 steps. The SHAKE algorithm was used to restrain hydrogen atom movement. The timestep was 1 fs and the run lasted 30 000 steps (30 ps).

***In vacuo* trimer - models based on a1 and a2 alignments:** It was suspected that mistakes in the alignment used for the early model (results not shown) were resulting in poor models, and thus poor MD. Therefore, new alignments were used as described above (2.3.2) to rebuild the trimers on three different templates (the alignment used for the early model was the same as a1, however only included the human template). Also, this time the effect of imposing symmetry on internal restraints was monitored.

The system was minimised and equilibrated as described above (2.3.3.1), except that the equilibration time varied between 30 000 and 90 000 steps. The 30 000 step runs were for models based on the a1 alignment with symmetry imposed, and for the FUGUE based a2 alignment without symmetry. The 90 000 step runs were for models based on the a2 alignment with symmetry imposed, using separate runs with and without the active site Mg^{2+} atoms.

Sampling: For the production run the system was then simulated for 1 to 10 ns (10 000 000 steps) at constant energy. Sampling was only pursued for systems equilibrated for 90 000 steps.

Hardware: CHARMM was run on a cluster of 64 Pentium IV processors running Gentoo Linux, with GigaBit Ethernet interconnect at the University of Pretoria.

2.3.3.2 *In vacuo* simulations with NAMD

Although CHARMM was initially used for molecular dynamics, it was found to be too slow on the hardware that was initially available (Pentium IV Beowulf cluster with GigaBit Ethernet). NAMD 2.6 was therefore used instead because of being faster and scaling better with more processors.

Long range interactions: A cut-off of 12 Å was used for electrostatic and van der Waals interactions. These interactions were gradually tapered to 0 from 8 to 12 Å (switching function). All pairwise interactions more than 3 covalent bonds apart were included, as for the CHARMM simulations. The same non-bonded parameters were used for minimisation and dynamics.

Minimisation: For adding hydrogens CHARMM was still used. Hydrogens were added as for the CHARMM simulations: all His residues were again N_δ protonated, except for 193 and 218 which were N_ϵ protonated. The system was first minimised for 450 steps.

Dynamics: The system was heated from 60 K to 310 K in 10 K increments every 500 steps with velocity reassignment. Equilibration was for a total of 200 000 (200 ps) steps of 1 fs each. For the production run the system simulated at constant energy for 1.5 ns (with Mg^{2+}) or 12 ns (without Mg^{2+}).

Hardware: NAMD was run on a cluster of 64 Pentium IV processors running Gentoo Linux, with GigaBit Ethernet interconnect at the University of Pretoria.

2.3.3.3 Solvated simulations with NAMD - NP sampling

The following parameters were used to simulate the trimer in an isobaric (NP) ensemble.

Long range interactions: Non-bonded interactions were shifted to zero from 10 Å to a cut-off of 12 Å. All non-bonded interactions connected by more than four covalent bonds were included. The trimer complex was first minimised *in vacuo* for 450 steps.

All steps including the complete system with solvent and ions were simulated with periodic boundary conditions using Particle Mesh Ewald (PME) Sums for the electrostatic calculations. A typical cell size was about 118x Å × 113y Å × 79z Å. A PME sub-angstrom grid size of 120 × 120 × 90 points was used.

System setup: Hydrogen atoms were added automatically using CHARMM 32B1. The histidine protonation scheme adopted is based on the requirements for catalytic activity and co-ordination of metal atoms in the active site. Thus His 193 and His 218 which were protonated on N_ϵ, and His 233 which was protonated on both N-atoms. The double protonation of His 233 is in accordance with the high resolution (1.29 Å) crystal structure of human arginase I (PDB id: 2AEB, Di Costanzo *et al.*, 2005), for which hydrogen positions were also determined, as well as previous speculation concerning activity. All remaining Histidines were protonated in the N_δ atom. Glu, Asp and Lys residues were charged. Since the current CHARMM forcefields are not parametrised for Mn^{2+}, the Mg^{2+} ion was used instead. It was thus assumed that any effects of the metal on trimer formation were largely electrostatic in nature.

The system was explicitly solvated using the SOLVATE plugin of VMD using the TIP3 water model. The protein was padded with solvent for 12 Å in the x- and y-directions (xy being coplanar with the trimer), and 10 Å in the z-direction. NaCl counter ions were added with the AUTOIONIZE from VMD plugin to an ionic strength of 50 mM with charge balancing to create a net system charge of zero. The system with metals (Mg^{2+}) included 14 Na^+ ions and 11 Cl^- ions, while the system without metal finally contained 20 Na^+ ions and 5 Cl^--ions. Assuming a total protein volume of 20% of the cell size (118 Å × 113 Å × 79 Å), 25 ions gives a concentration of 50.3 mM.

$$\frac{25 \text{ ions}}{0.8 \times 117 \times 10^{-10}\, m \times 112 \times 10^{-10}\, m \times 78.7 \times 10^{-10}\, m \times 1000 \frac{l}{m^3}} \times \frac{1\, mol}{6.02214179 \times 10^{23} \text{ ions}} = 50.3\, mM$$

It was realised later after the lengthy production runs that the program uses the concentration parameter to specify total ion concentration, and therefore the amount of each ion is less. However, 50 mM is a standard ionic concentration used in the molecular dynamics literature.

Heating: The solvent of the entire system was first minimised for 2000 steps, followed by the solvent and non-backbone atoms (backbone = C,N,O,C_α) for another 2000 steps. The entire system was then heated for 20 000 steps (20 ps) from 60 K at 10 K increments every 500 steps using velocity reassignment. All atoms were then minimised for 2000 steps, followed by another heating step as described, but for 200 ps. During heating timesteps were 1 fs and a Langevin piston (piston period 100 fs, piston decay 50 fs) was used to maintain pressure at 1 atm.

Equilibration and sampling: A Langevin piston was used to maintain pressure at 1 atm (NP ensemble) with a piston period of 200 fs and a piston decay of 100 fs, whereas temperature was left to fluctuate. Total run lengths for this stage were 20 ns.

Hardware: NAMD was run on a cluster of 64 Pentium IV processors running Gentoo Linux, with GigaBit Ethernet interconnect at the University of Pretoria.

2.3.3.4 Solvated simulations with NAMD - NPT sampling

Long range interactions and system setup were the same as for the NP simulations. The following parameters were used to simulate the trimer in an isobaric-isothermal ensemble (NPT).

Heating: Two different heating protocols were used. The first protocol (NPT_1) is the same as that used for the NP ensemble. In the second protocol (NPT_2) all steps were the same as for the first, except that during the first heating only solvent atoms were allowed to move. During heating timesteps were 1 fs and a Langevin piston (piston period 100 fs, piston decay 50 fs) was used to maintain pressure at 1 atm.

It was initially intended to use the second protocol, but the first was used instead because of a typographical mistake in the input. During heating and minimisation, the general procedure is to gradually prepare the system for long dynamics runs at high temperature, since starting a simulation at the desired temperature directly, can result in numerical errors. For this, a number of different protocols are possible, thus it is not expected that using this slightly different protocol will produce vastly differing results. After gaining access to faster hardware (described below), it became feasible to repeat the molecular dynamics run using the slightly altered protocol.

Equilibration and sampling: A Langevin piston was used to maintain pressure at 1 atm with a piston period of 200 fs and a piston decay of 100 fs, and Langevin dynamics was used

to also maintain temperature at 310 K (NPT ensemble) with a damping constant of 5 ps^{-1}. Total run lengths for this stage were in the order of 20 - 50 ns.

Hardware: A number of Linux clusters were used for molecular dynamics simulations. These include a 64 processor Gentoo Linux cluster of Pentium IV processors (University of Pretoria), Clusters of Xeon, Opteron or Itanium2 processors (BMIC, Meraka Institute, CSIR) running Scientific Linux. In both cases the interconnect comprises GigaBit Ethernet. Temporary access during the setup phases was also granted to The Centre for High Performance Computing (CHPC, Cape Town, South Africa). The CHPC hardware comprises Multicore Opteron processors with InfiniBand interconnect.

2.3.4 Analysis

VMD (Humphrey *et al.*, 1996) was used for analysis of the molecular dynamics data. VMD and PYMOL (Delano, 2002) were both used for visualisation. The STRIDE program as shipped with VMD was used to assign secondary structure. The MULTISEQ plugin was used to visualise secondary structure alignments. Salt-bridges between arginine and glutamate residues were measured between the arginine C_ζ and glutamate C_δ, atoms respectively. A distance of about 4 Å corresponds to the typical distance of 2.8 Å between the hydrogen bond donor (PDB naming: NH_x) and acceptor atoms (PDB naming: OE_x) respectively.

2.3.5 Site-directed mutagenesis (IBM)

The mutations *Pf* Arg-Glu 295 Ala, *Pf* Arg-Glu 295 Arg, *Pf* Arg-Arg 404 Ala as well as the double mutation *Pf* Arg-Glu 259 Ala/Arg 404 Ala were introduced into the recombinant *P. falciparum* arginase by site-directed mutagenesis; activity and oligomeric status were detected as described in Müller *et al.* (2005). This work was performed by the collaborating group of Prof. Walter at the Biochemistry department of the Berhnard-Nocht Institute for Tropical Medicine, Hamburg, Germany.

2.3.6 Simulation of mutants

SCWRL3 (Canutescu *et al.*, 2003) was used to introduce mutations into the homology model prior to addition of hydrogens and molecular dynamics. The same protocol was then followed as for the NPT 2 simulations.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

2.4 Results and discussion

2.4.1 Sequence alignment and homology modeling

2.4.1.1 *Plasmodium*-specific inserts

Searching the online *Plasmodium* genome resource, PlasmoDB (Bahl *et al.*, 2003), revealed the arginase sequences for *P. vivax*, *P. yoelii*, *P. knowlesi* and *P. berghei*, in addition to the previously characterised *PfArg* (Müller *et al.*, 2005). From the automated alignments, two *Plasmodium*-specific inserts were identified. Alignments generated from CLUSTALW and FUGUE were used as references during manual adjustment of the alignments used for modeling. In both reference cases, the positions of both inserts do not differ markedly. According to the alignment ultimately used for modeling, the first insert runs from residues 77 to 151 (75 residues), and the second from residues 377 to 388 (12 residues). Insert 1 varies considerably in sequence and length between different *Plasmodium* species, ranging from approximately 100 residues in *P. vivax* to only 15 residues in *P. berghei*. It is predicted to lie between the second β-strand and second α-helix on the outer edge of the trimer (Fig. 2.8). In contrast, insert 2 is highly conserved in sequence and length in all *Plasmodium* species. Insert 2 is located between the last β-strand and the last α-helix. The sequence identity between the *P. falciparum* and templates was 35%, 30% and 27%, respectively for the bacterial, rat and human templates. A shift of one residue for insert 2 had significant effects on its conformation relative to the trimer. When modeling according to CLUSTALW based alignments, insert 2 was predicted to interact at the trimer interface. However, when using FUGUE based alignments insert 2 was predicted to fold back away from the trimer interface and interact with it's respective monomer (Fig. 2.8).

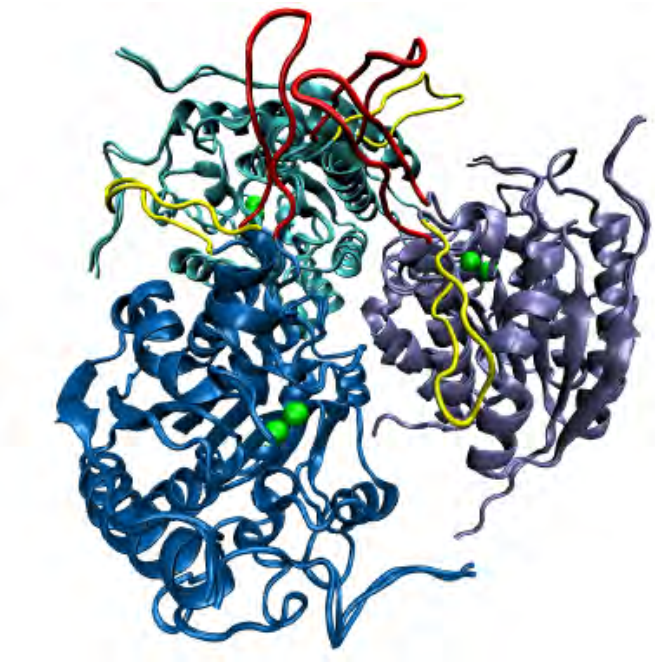


Figure 2.8: Effect of alignment on conformation of insert 2. When moving one residue upstream, the insert folds away from the trimer interface (yellow) compared to making contact (red). The active site Mg^{2+} atoms are indicated in green. Monomers are distinguished by different shades of blue.

Proteins from *Plasmodium* frequently have large inserts relative to sequences from homologues in other organisms (Aravind *et al.*, 2003). They are often characterised by low com-

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

plexity (Pizzi and Frontali, 2001; Xue and Forsdyke, 2003) and/or have a strong amino acid bias towards small and hydrophilic residues. The insert-originating mechanism remains unknown. Possible global functions of these inserts include nucleic acid level adaptations (Pizzi and Frontali, 2001; Xue and Forsdyke, 2003), immune evasion and protein-protein interactions (Aravind *et al.*, 2003), however these remain speculative. It has been demonstrated that interfering with these inserts can affect the protein, and thus they may have local functions relative to their enzymes (Yuvaniyama *et al.*, 2003; Birkholtz *et al.*, 2004; Jean *et al.*, 2005; Rattanachuen *et al.*, 2009). These inserts tend to align with non-core regions in homologues for which the structure is known. Most of insert 1 was left un-modelled (residues 81 to 147 removed) due to its length, leaving a small overhang at each end with respect to the templates used (Fig. 2.6-2.7). Insert 2, however, being much shorter was retained for *ab initio* modeling. A large number of sequences were used to generate multiple alignments used as references during the editing of alignments for modeling. While the reference alignments were often highly redundant, all sequences were retained in order to offset the presence of five *Plasmodium* sequences. These were included to aid in delineation of *Plasmodium*-specific inserts.

Because a small change in alignment had a substantial effect on insert 2, it is important to justify the choice of reference alignment. The FUGUE program makes use of environment-specific substitution tables and structure-dependent gap-penalties, and is thus generally expected to give a more accurate starting alignment for modeling purposes. Furthermore, the CLUSTALW alignment was also found to disturb important inter-monomer interactions. Preference was therefore given to the FUGUE reference when adjusting alignments used for modeling.

2.4.1.2 Active site and inter-monomer residue conservation

The model preserves standard active site residues observed in other arginase structures. All Mn^{2+} co-ordinating residues previously identified are conserved in the model (Section 2.4.3.6). The only substitutions are in second shell ligands when compared with the bacterial template, where Ser 176 and Glu 268 (*B. caldevelox*) are replaced by Asp 272 and Asp 365 (*P. falciparum*), respectively. Residues implicated in substrate binding are also highly conserved. There are no substitutions compared to the bacterial template and only one compared to the mammalian templates: Thr 135 (rat) is replaced by Ser 227 in the model. The high conservation of the active site suggests that inhibitors specific to the *P. falciparum* active site will be difficult to find. Thus, an alternative means of inhibition may be necessary if *P. falciparum* arginase is to be of potential therapeutic value. Therefore, attention was directed at the inter-monomer interactions.

2.4.1.3 Inter-monomer interactions

The main interaction for oligomerisation of arginase into the trimer was found to be a salt-bridge between Arg 255 and Glu 256 in adjacent monomers of human arginase I (Sabio *et al.*,

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

2001). This interaction is conserved in all arginase structures studied so far: Arg 249$_a$::Glu 250$_b$ (*B. caldevelox*), Arg 255$_a$::Glu 256$_b$ (human and rat arginase I), Arg 274$_a$::Glu 275$_b$ (human and rat arginase II). In *P. falciparum,* the corresponding interaction is formed by Arg 346$_a$::Glu 347$_b$.

Visual inspection of the models suggested a novel inter-monomer salt-bridge forms between Glu 295$_a$::Arg 404$_b$ (Fig. 2.9). In multiple sequence alignments, Glu 295 aligns with conserved acidic residues in the bacterial and mammalian templates. *P. falciparum* Glu 295 aligns with an Asp in mammals (human arginase II: Asp 223, rat arginase I: Asp 204), fungi and bacteria (*B. caldovelox* arginase: Asp 199). In the other *Plasmodium* species, Glu 295 aligns either with Asp (*P. yoelii* and *P. berghei*) or Glu (*P. knowlesi* and *P. vivax*). The only exception is in plants, where a serine residue is found instead (*Arabidopsis thaliana* arginase I: Ser 247). In the model, and the mammalian and bacterial templates this acidic residue forms interactions with the adjacent monomer via partner residues that do not align. In mammals, Asp 223/204$_a$ forms a salt bridge with Arg 308/327$_b$ (rat arginase I/human arginase II) instead. This salt-bridge nucleates considerable inter-monomer interactions, characterised by an S-shaped *C*-terminus (Kanyo *et al.*, 1996; Cama *et al.*, 2003a). The *Plasmodium* sequences lack such an extensive *C*-terminus and Arg 308 aligns with a hydrophobic Ile instead (408: *P. falciparum*, 368: *P. knowlesi*, 436: *P. vivax*, 353: *P. berghei*, 376: *P. yoelii*). In the bacterial structure the Asp 199 cognate forms an inter-monomer bridge with Glu 256 that is mediated either by urea or by free arginine, depending on the crystallisation conditions (Bewley *et al.*, 1999). The *Plasmodium* Arg salt-bridge partner to Glu 295 aligns with small and/or hydrophilic residues in other organisms (eg. Ser, Thr, Cys, Ala, Glu). The importance of the S-shaped tail is still in doubt since products truncated after Arg 308 can still form active trimers (Mora *et al.*, 2000). In bacterial structures, an interaction is formed with another acidic residue (Glu 256) that is mediated by either urea or free arginine. Finally, in the *P. falciparum* model Glu 295 is predicted to interact with Arg 404. Arg 404 doesn't align with Arg 308 or Glu 256. Thus, there appears to have been evolutionary pressure to establish a strong inter-monomer interaction in this region of the monomer-monomer interface. The differences between the *P. falciparum* model and templates suggest this salt bridge as a possibly unique interaction and was therefore subject to scrutiny using molecular dynamics and site-directed mutagenesis.

In order to determine other possible interactions the salt bridge analysis tool of VMD was used to search for all possible salt bridges in the protein, using co-ordinates prior to sampling. All salt bridges with a hydrogen-bond donor/acceptor distance less than 3.2 Å were identified. Only one other interaction between adjacent monomers was found, between Glu 400 and Lys 340. However, this interaction was not stable during molecular dynamics. This was observed both with and without Mg^{2+} (results not shown). Thus, this interaction is likely to be only of secondary importance in maintaining quaternary structure, and attention was focused on the Glu 295::Arg 404 interaction instead.

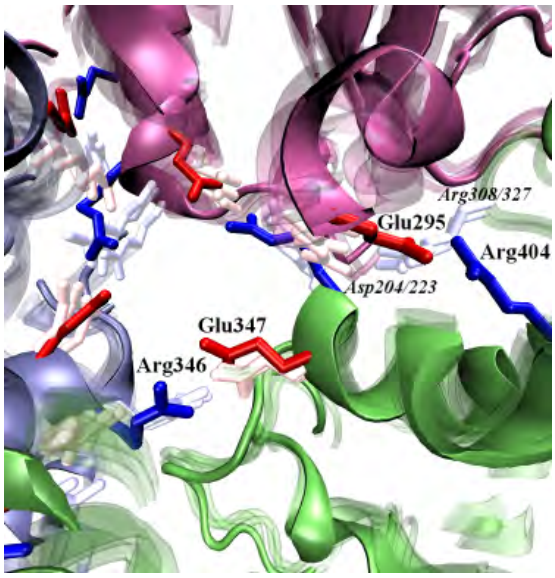


Figure 2.9: Inter-monomer salt bridges in *Pf*Arg. The conserved inter-monomer salt-bridge at Arg 346$_a$::Glu 347$_b$ and the non-conserved inter-monomer salt-bridge at Glu 295$_a$::Arg 404$_b$ are both indicated. Monomers are in green, light-blue and mauve. The rat arginase 1 and human arginase II templates are shown in transparency, and their respective salt-bridges indicated in lighter shades of red (acidic) and blue (basic). Template residues are in italics (*rat/human*). An interactive 3D summary of the various inter-monomer interactions is given in Fig. A.1

2.4.2 Initial simulations

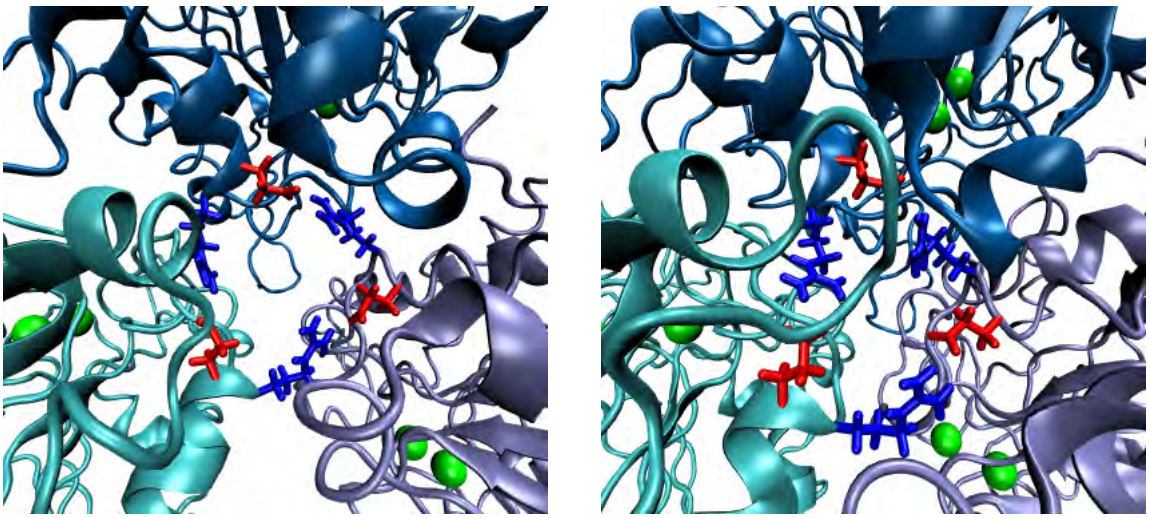
2.4.2.1 Simulation times with CHARMM

Simulating the complete trimer with CHARMM was found to be quite slow, even in parallel. Using 4 CPUs, heating the shell solvated system (2.3.3.1) took approximately 14 hours for 30 000 steps (± 55 000 atoms). This amounts to 1.6 s of wall-clock time per 1 fs timestep in the simulation. The equivalent simulation of the trimer without solvent (± 16800 atoms 2.3.3.1, 2.3.3.1) took approximately 1.35 hours with 16 CPUs (0.18 seconds per timestep (1 fs)), 1.45 hours with 12 CPUs (0.3 seconds per timestep (1 fs)) and 2.5 hours with 4 CPUs (1.44 seconds per timestep (1 fs)). Simulating models based on the a2 alignment (± 15900 atoms) on 8 CPUs gave a time of about 0.23 s per timestep (1 fs). By comparison, simulating ± 96 000 atoms with NAMD using 4 CPUs took 1.4 s per timestep (2 fs). Thus, with CHARMM simulating comparatively small systems took a relatively long time. While the shell solvated system is relatively small, it will not remain intact. The solvation shell attempt to lower surface tension over time and this converge on a spherical droplet modified by the presence of the protein solute. Furthermore, individual solvent molecules will drift away from the system due to random chance. Thus, it is not suitable for accurate simulations with explicit solvent. Instead, explicit solvation requires appropriate boundary conditions that will ensure the protein surface is in contact with solvent. Because of the slow simulation times that would have accompanied a fully solvated protein box, explicit solvation was abandoned for simulations in CHARMM.

2.4.2.2 Trimer interface integrity with CHARMM

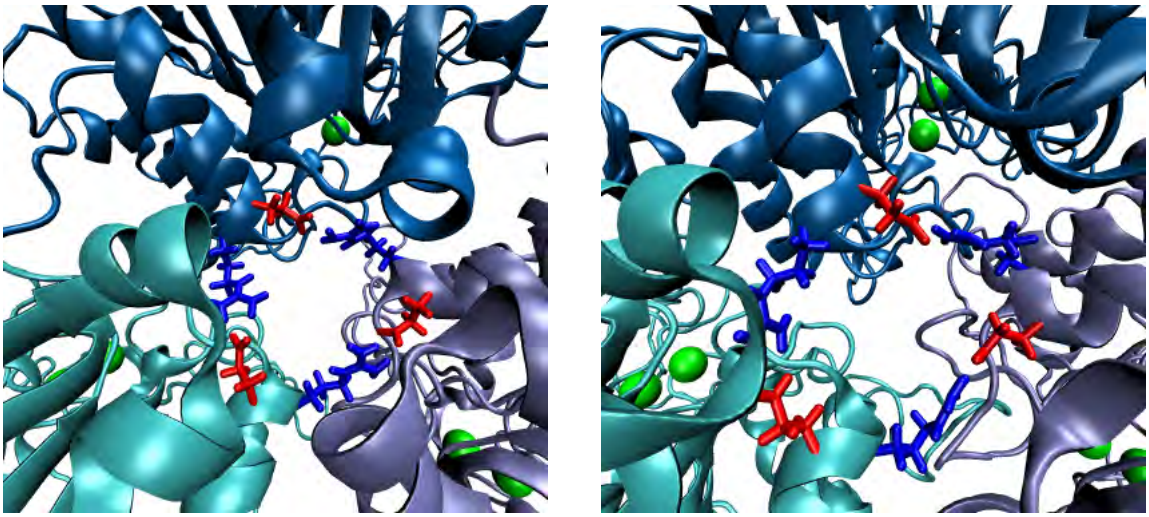
The salt-bridge at Arg 346$_a$::Glu 347$_b$ is conserved in all known arginase structures. It is required for maintaining the trimer complex, and hence it's integrity during MD simulations was closely monitored. During *in vacuo* equilibration of the early model this interaction was disturbed or poorly maintained (Fig. 2.10). The trimer interface was also observed

to collapse during *in vacuo* simulations. Imposing symmetry on the internal co-ordinates was observed to improve this slightly, but there was still collapse of the trimer interface (Fig. 2.11). Using models built on the a2 alignment without symmetry did not improve the situation at the trimer interface either (Fig. 2.12). Combining both the use of the a2 alignment and symmetry was observed to have less of a detrimental effect on the trimer interface, even after 90 000 steps (90 ps) for systems both with and without Mg^{2+}. Therefore, these simulations were extended for production/sampling runs. However, after 10 ns the same kind of collapse at the trimer interface was observed (Fig. 2.13).



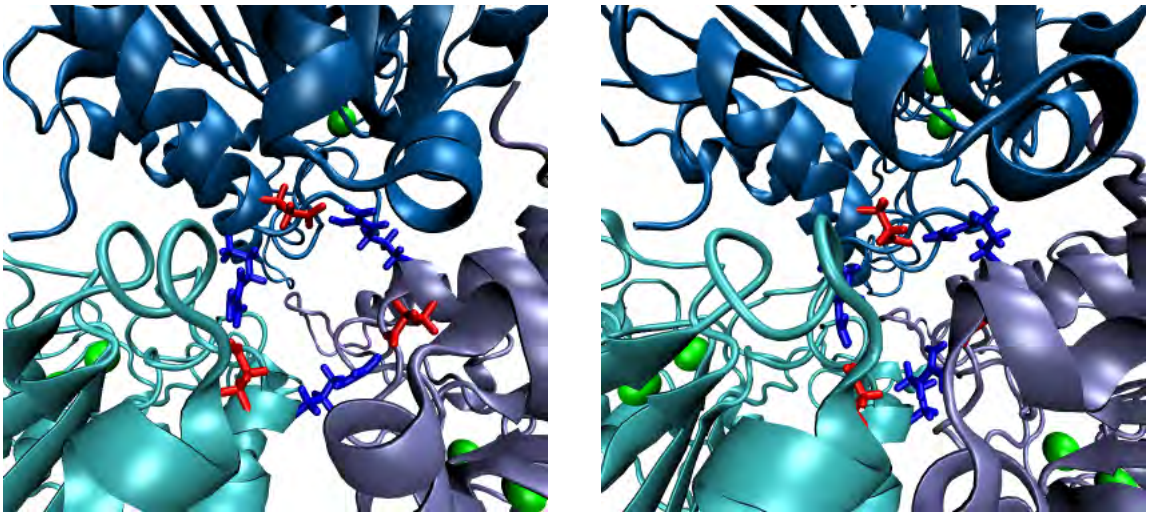
(a) Before equilibration (b) After equilibration

Figure 2.10: *In vacuo* equilibration in CHARMM with an early model. During equilibration the trimer interface usually collapses, which disturbs the conserved interaction at Arg 346_a::Glu 347_b Highlighted residues (blue: Arg, red: Glu) are shown as stick representations. The different chains are indicated in different shades of blue. The active site metals are depicted as green spheres.



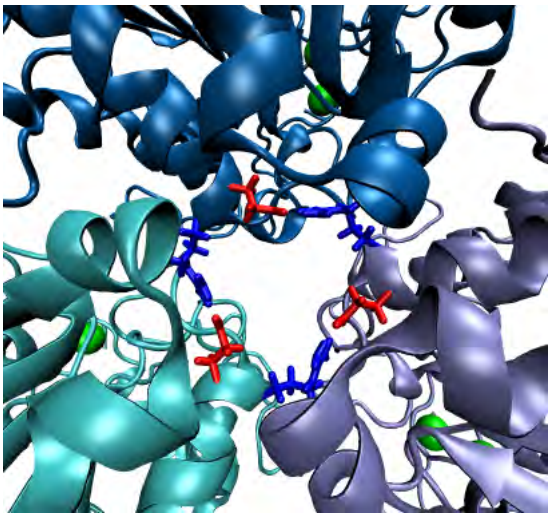
(a) Before equilibration (b) After equilibration

Figure 2.11: *In vacuo* equilibration in CHARMM with an alignment 1 (a1) derived model including symmetry. During equilibration the trimer interface is still observed to collapse, however less so than with the MD on the early model. The conserved interaction between Arg 346_a::Glu 347_b is indicated. Highlighted residues (blue: Arg, red: Glu) are shown as stick representations. The different chains are indicated in different shades of blue. The active site metals are depicted as green spheres. The active site metals are depicted as green spheres.

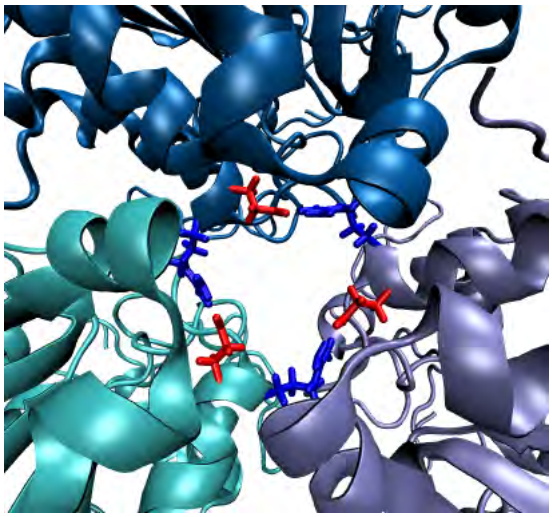


(a) Before equilibration (b) After equilibration

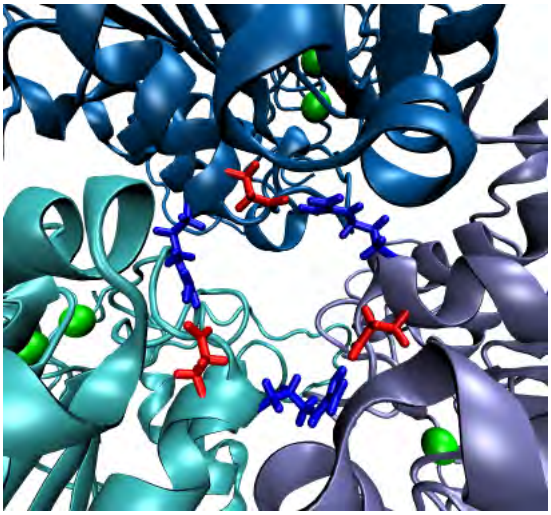
Figure 2.12: *In vacuo* equilibration in CHARMM with alignment 2 (a2) derived model excluding symmetry. During equilibration the trimer interface collapse is again observed. The conserved interaction between Arg 346_a::Glu 347_b is indicated. Highlighted residues (blue: Arg, red: Glu) are shown as stick representations. The different chains are indicated in different shades of blue. The active site metals are depicted as green spheres.



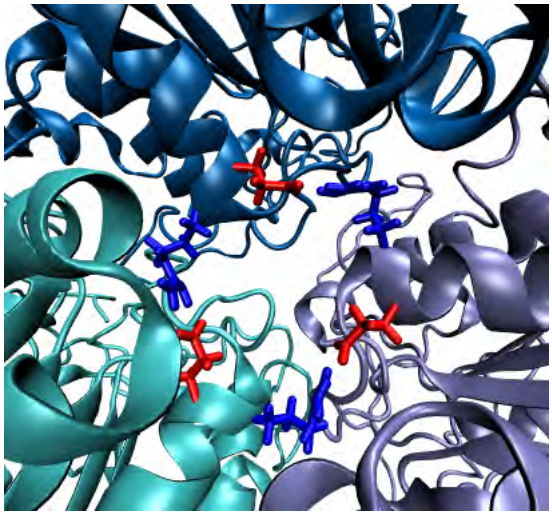
(a) Before equilibration (+ Mg^{2+})



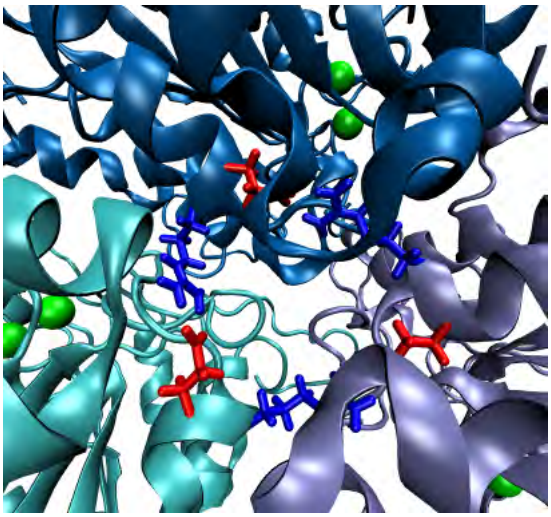
(b) Before equilibration (- Mg^{2+})



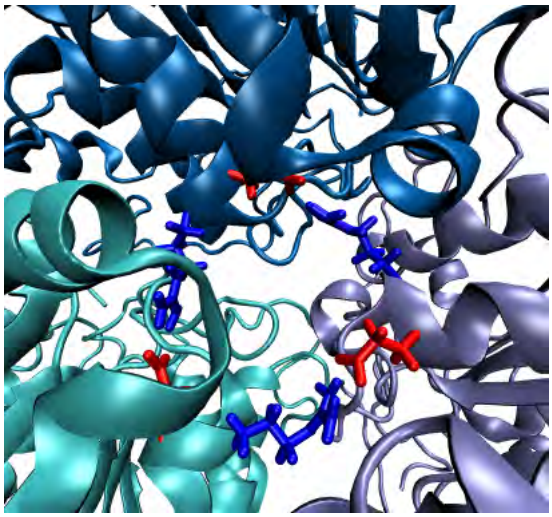
(c) After equilibration (+ Mg^{2+})



(d) After equilibration (- Mg^{2+})



(e) After 10 ns sampling (+ Mg^{2+})



(f) After 10 ns sampling (- Mg^{2+})

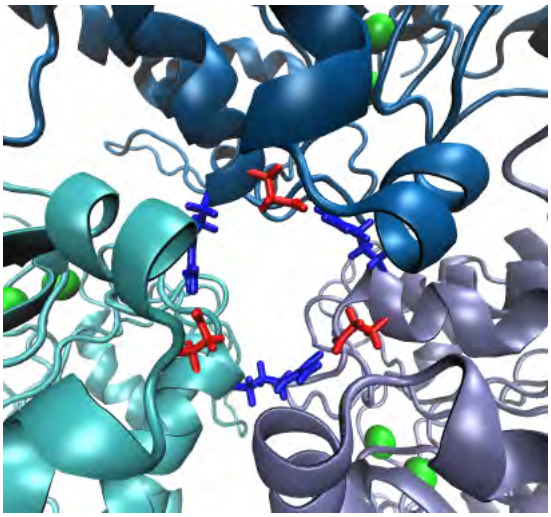
Figure 2.13: *In vacuo* equilibration with alignment 2 (a2) derived including symmetry. During equilibration the trimer interface collapse is again observed and worsens during sampling. The conserved interaction at Arg 346_a::Glu 347_b is indicated. Highlighted residues (blue: Arg, red: Glu) are shown as stick representations. The different chains are indicated in different shades of blue. The active site metals are depicted as green spheres.

2.4.2.3 *In vacuo* simulations with NAMD

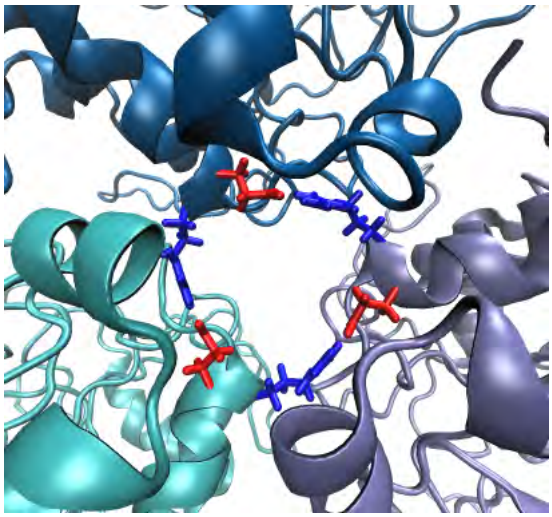
Similar behaviour was observed as for the CHARMM simulations. At the end of equilibration one of the Arg 346_a::Glu 347_b salt-bridges was broken in the simulation with Mg^{2+}, while all three salt-bridges were still intact in the system without Mg^{2+}. Collapse of the central pore was also observed in both systems. After 1.5 ns of the production run this was more pronounced in both cases and all three salt-bridges were broken. There was also more visible distortion and loss of secondary structure in the simulation without Mg^{2+} (Fig. 2.14). By the end of 12 ns of the production run (executed for the non-Mg^{2+}simulation only) the destabilisation of the protein had increased further.

Although both systems with and without Mg^{2+} showed considerable distortion during molecular dynamics, there was more distortion in the non-metal system. This is reflected in the Root Mean Square Deviation (RMSD) of the positions of the $C_\alpha$atoms during the production run. In the system with Mg^{2+}removed, the RMSD compared to the first frame shows a greater increase (Fig. 2.15). This indicates that there is more structural distortion of the trimer complex with the active site metal removed.

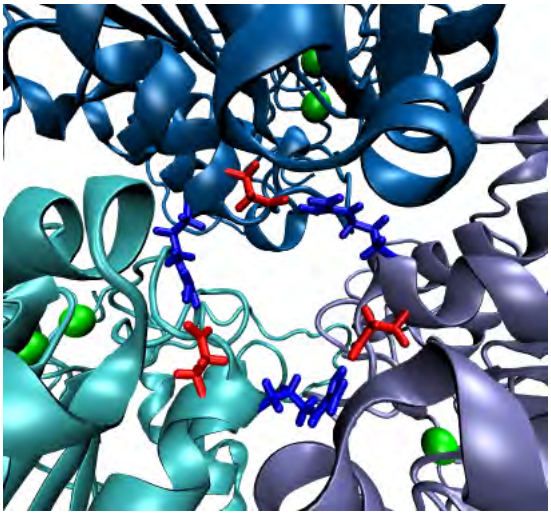
Simulating *in vacuo* with NAMD was judged to be inadequate for investigating the structural metal dependency. Firstly, because the conserved interaction at Arg 346_a::Glu 347_b was not stable in the simulation with Mg^{2+}, and secondly because of the collapse of the central pore of the trimer interface after only a short time of simulation. These results further confirmed the necessity of having solvent in the central pore to prevent collapse.



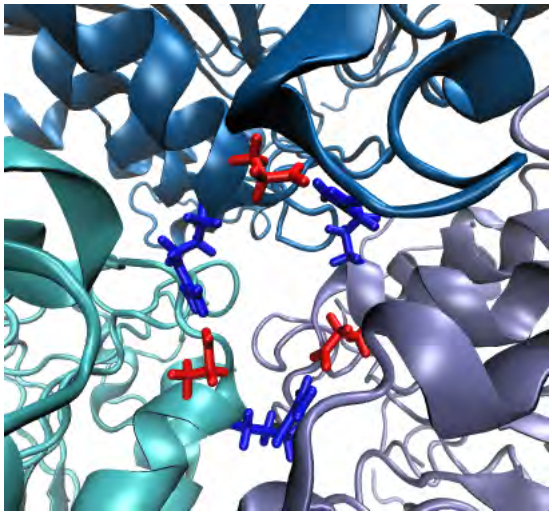
(a) Before equilibration (+ Mg^{2+})



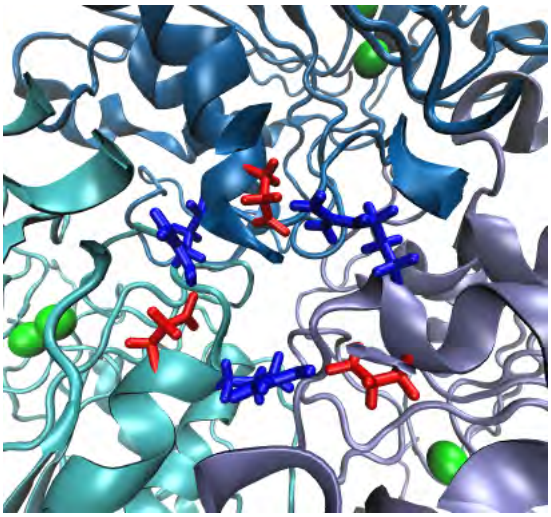
(b) Before equilibration (- Mg^{2+})



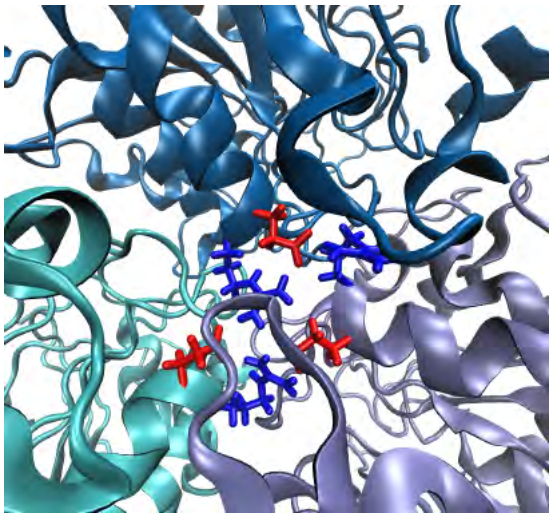
(c) After equilibration (+ Mg^{2+})



(d) After equilibration (- Mg^{2+})



(e) After 1.5 ns sampling (+ Mg^{2+})



(f) After 1.5 ns sampling (- Mg^{2+})

Figure 2.14: *In vacuo* simulations with NAMD - a2 alignment (with symmetry and metals). During equilibration the trimer interface collapse is again observed. The conserved interaction at Arg 346_a::Glu 347_b is indicated. Highlighted residues (blue: Arg, red: Glu) are shown as stick representations. The different chains are indicated in different shades of blue. The active site metals are depicted as green spheres.

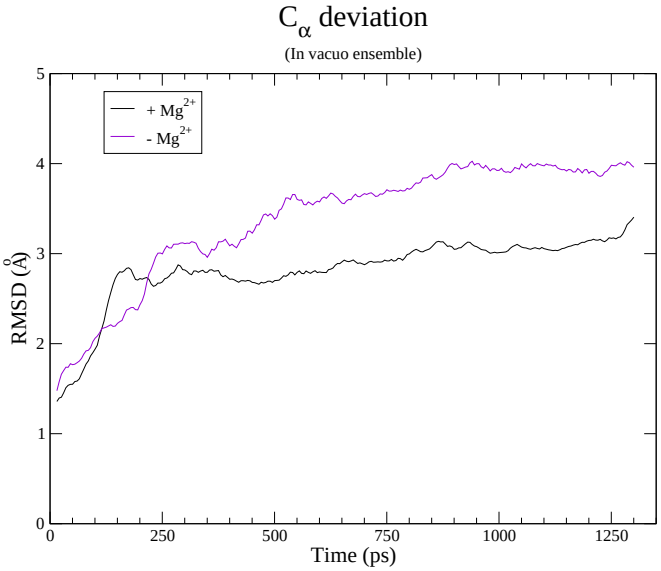


Figure 2.15: C_α RMSD during the production run of the *in vacuo* ensemble with and without Mg^{2+}. Removing Mg^{2+} causes a greater increase in RMSD as sampling progresses. The plots were averaged using a sliding window of 50 frames (500 fs).

2.4.2.4 Conclusions

In general it was found that *in vacuo* simulations in CHARMM were not suitable for investigating the structural metal dependency. In all simulations the trimer interface was observed to collapse, often destroying the integrity of the Arg 346_a::Glu 347_b salt-bridge which is required for trimer integrity. For reliable simulations this interaction should remain intact in conditions not expected to disturb the trimer. The interface collapse is likely due to the absence of any solvent species occupying the centre of the trimer which would help prevent collapse through steric interactions. The presence of explicit solvent would also shield strong non-bonded interactions across the central pore. Furthermore, the CHARMM simulations were found to be fairly slow, with poor scaling when running on larger numbers of CPUs. During this, NAMD was also tested which is known to scale well on the kind of hardware that was available. The better performance of NAMD is probably due to the optimised message-passing and spatial decomposition algorithms for commodity-hardware based clusters (Phillips *et al.*, 2005). It was found that fully solvated NAMD simulations (± 96 000 atoms), executed at approximately the same speed as un-solvated CHARMM simulations (± 16 000 atoms). However, as with CHARMM, the integrity of the trimer interface during *in vacuo* simulations was also not maintained with NAMD. Therefore, because the execution speed was the same and including solvent is more biologically realistic, it was decided to do all further simulations with NAMD and explicit solvent.

2.4.3 Solvated simulations

2.4.3.1 Introduction of chain break for insert1

Before proceeding with detailed analysis, it was necessary to ensure that gross changes to the protein necessitated by the modeling difficulties would not compromise the interpretation of results. The omission of insert 1 is potentially problematic, in that it introduces an unnatural

chain-break and therefore potential instability into the protein. The deletion of the insert creates a protein fragment on the outer edge of the trimer complex that does not interact extensively with any neighbouring monomers, and makes most of its contacts with it's own monomer. The insert was too long to attempt any modeling of this using the protocols described. The gap was left un-ligated, and apart from the loss of some secondary structures in this region (described below), this fragment was stable for up to 50 ns of simulation and remained in contact with the rest of the protein. The deletion of insert 1 for modeling did not adversely affect the stability of the model: while potential problems by introducing a chain break could have been avoided by ligating the ends of the gap, this would also be unnatural. Because the fragment was apparently stable and closing the gap un-ligated is less parsimonious, the break was left in.

2.4.3.2 RMSD

Protein stability was monitored by the change in C_α RMSD during the sampling runs. C_α RMSD was compared with the first frame during the equilibration and sampling runs. In all solvated simulations there is an increase in RMSD which equilibrates after about 10 ns for the NP simulation, and between 20 and 30 ns for the NPT simulations. In two out of three (NP: Fig. 2.16 , NPT_2: Fig: 2.18), the C_α RMSD equilibrates at 1 Å more than in the system with Mg^{2+}, similar to what was observed for the *in vacuo* simulation. In the NPT_1 simulation the situation is more variable, with the RMSD of -Mg^{2+} less than that of the Mg^{2+} system between 10 and 30 ns, but larger for the first 10 and last 5 ns out of 55 ns of simulation (Fig. 2.17). The equilibration of C_α RMSD at a larger distance for the -Mg^{2+} systems suggests that removing the active site metals has a detrimental effect on protein stability. It was previously reported that removing Mn^{2+}, either by dialysis and chelation with EDTA, or by mutagenesis of Mn^{2+} co-ordinating residues in the active site, not only abolished enzyme activity but also promoted dissociation of the trimer into monomers. Reformation of the trimer could be obtained by addition of Mn^{2+} (Müller *et al.*, 2005).

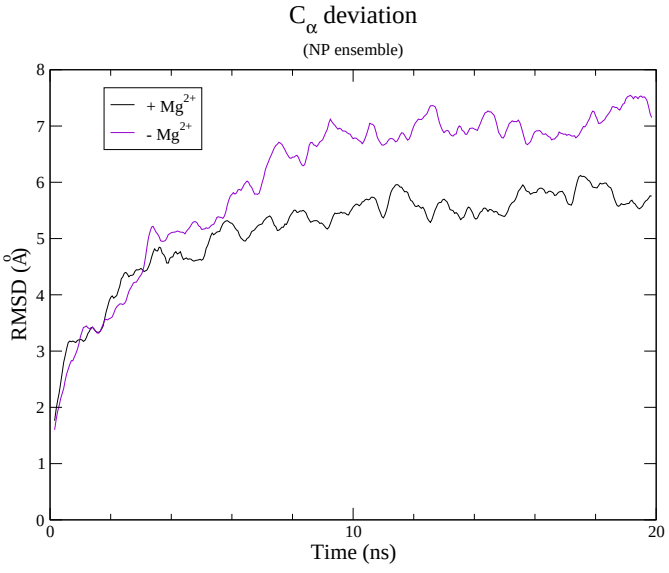


Figure 2.16: C_α RMSD during the production run of the NP with (+) and without (-) Mg^{2+}. Removing Mg^{2+} causes a greater increase in RMSD as sampling progresses. The plots were averaged using a sliding window of 500 frames (250 fs).

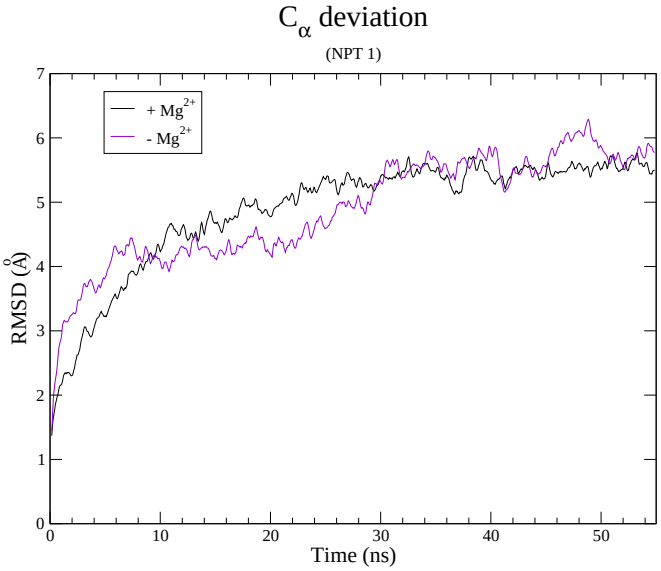


Figure 2.17: C_α RMSD during the production run of the NPT_1 with (+) and without (-) Mg^{2+}. There was no noticable difference with and without Mg^{2+} . The plots were averaged using a sliding window of 500 frames (250 fs).

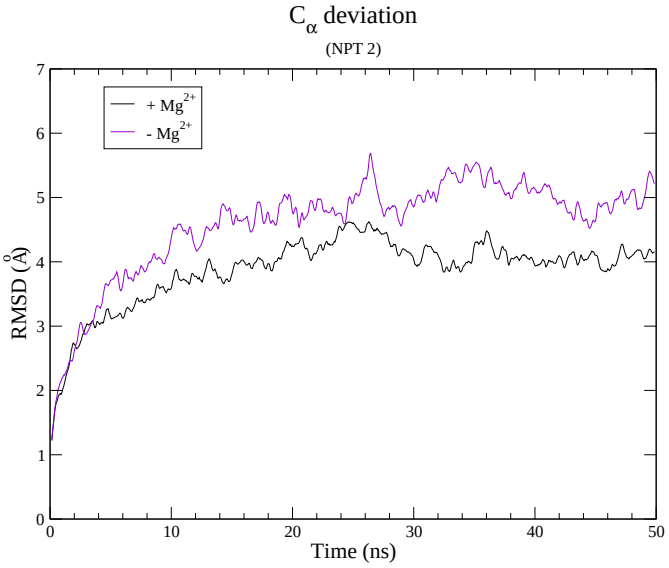


Figure 2.18: C_α RMSD during the production run of the NPT_2 with (+) and without (-) Mg^{2+}. Removing Mg^{2+} causes a greater increase in RMSD as sampling progresses. The plots were averaged using a sliding window of 500 frames (250 fs).

The RMSD of the mutant simulations was monitored as described above for the Mg^{2+} and Mg^{2+}-free simulations. In all mutants a similarly greater increase is observed compared to the wild type with Mg^{2+} (Fig. 2.19). The greatest increase occurs in the *Pf* Arg-Glu 347 Gln mutant, which equilibrates at ± 2.5 Å more than the wild type/Mg^{2+} by 40 ns. The increase is also the most rapid for *Pf* Arg-Glu 347 Gln mutant at ± 2 Å more by 10 ns. The next largest effect is observed for the Glu 295 Arg mutant, which largely mirrors the removal of Mg^{2+}, with a similar increase by 6 ns as for the Glu 347 Gln mutant. By 40 ns the RMSD for *Pf* Arg-Glu 295 Arg is 0.8 Å larger than wild type/Mg^{2+}. The *Pf* Arg-Glu 295 Ala, *Pf* Arg-Arg 404 Ala and *Pf* Arg-Glu 295 Ala/Arg 404 Ala mutants display a similar increase in RMSD by 40 ns, between 0.5-0.8 Å larger than for Mg^{2+}.

2.4.3.3 Preservation of secondary structure

The effect of removing the metal on conservation of secondary structure (Fig. 2.20) during sampling was also monitored. In general, a greater loss of secondary structural integrity was observed for non-metal systems in the NP and NPT_1 and NPT_2 ensembles (Table 2.3).

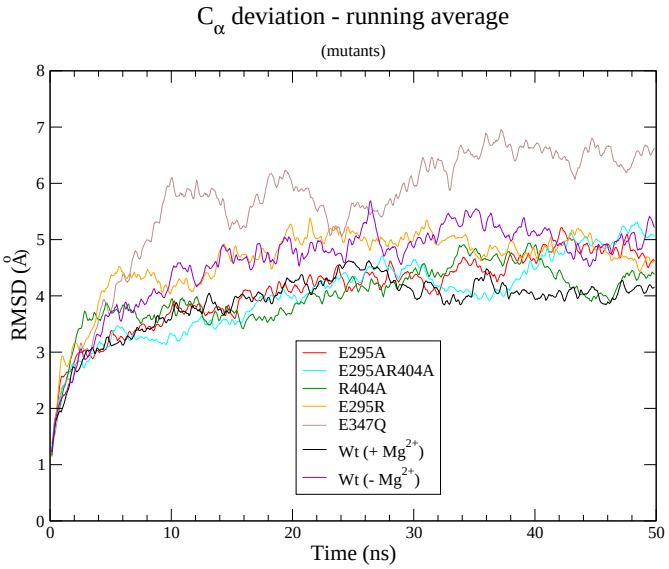


Figure 2.19: C_α RMSD of E295A, E295R, E347Q, R404A and E295AR404A mutants compared to wild type with Mg^{2+}. Introducing mutations into the inter-monomer salt-bridges causes a similar increases in equilibrium RMSD compared to removing Mg^{2+}. The plots were averaged using a sliding window of 500 frames (250 fs).

Combining all the simulations gives a total of 9 simulations for the monomer in each condition ($\pm Mg^{2+}$), which can be used to observe any general loss of secondary structure. However, these data are not sufficient to determine possible co-operative effects between monomers. In the absence of Mg^{2+}, complete loss of larger secondary structure is observed for few elements. The monomers/chains are arbitrarily designated A, B and C. In chain B of the NP simulation and chain C of NPT_1, the second β-strand is lost. The second α-helix of chain C in NPT_2 is also lost. For some of the smaller secondary structure elements ($\leq$4 residues) there is complete loss of structure for both conditions. This is observed for $\beta 3$, $\beta 4$, $\alpha 6$, $\alpha 10$ and $\alpha 12$. All of these secondary structural elements align with similar elements in all of the templates. The greater loss of secondary structure without $Mg^{2+}$is expected from the greater increase in RMSD. However, from this data there is no obvious protein region that suggests further investigation with regards to disturbing arginase activity.

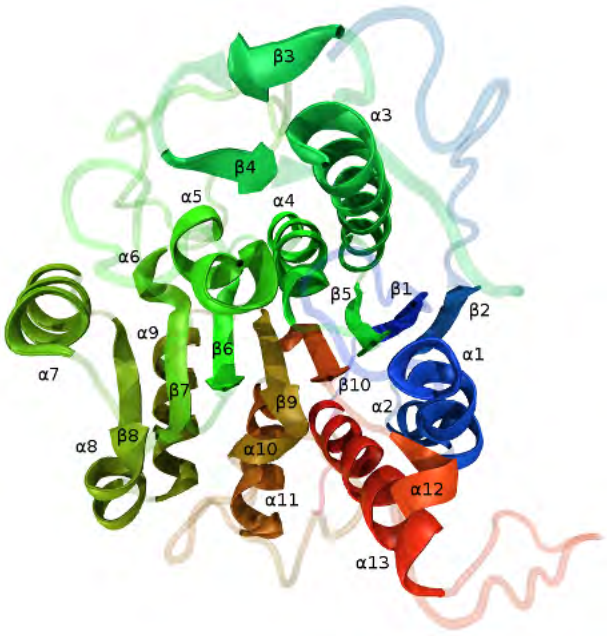


Figure 2.20: Tertiary structure of the *Plasmodium* arginase model. Secondary structure elements (αx for strands, βx for helices) are indicated. Increasing sequence position is indicated by a bluegreenred colour gradient. Loops and coiled regions are transparent for clarity.

Table 2.3: Preservation (+) and loss (-) of secondary structure during molecular dynamics. Each string approximately corresponds to the length of the secondary structure element (β-strands: 3-4 residues, $\alpha/3_{10}$-helices: 4-5 residues). The sum total of symbols is 450. A greater loss of secondary structure is observed upon removing Mg^{2+}: lost structure in simulations is represented by 48 and 65 "-" symbols in $+Mg^{2+}$and $-Mg^{2+}$simulations, respectively.

Condition	Simulation	Chain	$\beta1$	$\alpha1$	$\alpha2$	$\beta2$	$\beta3$	$\beta4$	$\alpha3$	$\beta5$	$\alpha4$	$\beta6$	$\alpha5$	$\alpha6$
$+Mg^{2+}$	NP	A	++	+++	-++	++	-	-	++++	++	+++	++	-+	-
		B	+-	+++	+++	+-	-	-	-+++	++	+++	++	-+	+
		C	++	+++	+++	++	-	-	++++	++	+++	++	-+	-
	NPT_1	A	++	-++	+++	++	-	-	++++	++	+++	++	-+	+
		B	++	+++	+++	++	-	-	-+++	++	+++	++	-+	+
		C	+-	-++	+++	+-	-	-	++++	++	+++	++	-+	-
	NPT_2	A	+-	+++	++-	+-	-	-	++++	++	+++	++	++	+
		B	+-	-++	+++	+-	-	-	++++	+-	+++	++	-+	-
		C	+-	+++	+++	+-	+	+	++++	++	+++	++	-+	+
		$\frac{-}{total}$	$\frac{5}{18}$	$\frac{3}{17}$	$\frac{2}{27}$	$\frac{5}{18}$	$\frac{8}{9}$	$\frac{8}{9}$	$\frac{2}{36}$	$\frac{1}{18}$	$\frac{0}{27}$	$\frac{0}{18}$	$\frac{8}{18}$	$\frac{4}{9}$
$-Mg^{2+}$	NP	A	+-	+++	+++	+-	-	-	++++	++	+++	++	++	+
		B	+-	+++	+++	--	-	-	--++	++	+++	++	++	-
		C	++	+++	+++	++	-	-	+-++	+-	+++	++	-+	-
	NPT_1	A	+-	-++	-++	+-	-	-	++++	++	+++	++	++	+
		B	+-	-++	+++	+-	-	-	++++	++	+++	++	++	+
		C	+-	+++	+++	--	-	-	-+++	++	+++	++	-+	-
	NPT_2	A	+-	+++	+++	+-	-	-	--++	+-	+++	++	-+	+
		B	+-	+++	+++	+-	-	-	++++	+-	+++	++	-+	+
		C	+-	+++	---	+-	+	+	++++	++	+++	++	-+	+
		$\frac{-}{total}$	$\frac{8}{18}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{10}{18}$	$\frac{8}{9}$	$\frac{8}{9}$	$\frac{6}{36}$	$\frac{3}{18}$	$\frac{0}{27}$	$\frac{0}{18}$	$\frac{5}{18}$	$\frac{3}{9}$

Table 2.3: Table 2.3 continued...

Condition	Simulation	Chain	β7	α7	β8	α8	α9	β9	α10	α11	β10	α12	α13
$+Mg^{2+}$	NP	A	++	++	++	++	+++	++	-	+++	++	-	-++
		B	++	++	++	++	+++	++	+	+++	++	+	-++
		C	++	++	++	++	+++	++	+	+++	++	-	-++
	NPT_1	A	+-	++	++	++	+++	++	+	+++	++	-	+++
		B	++	++	++	++	+++	++	-	+++	++	+	-++
		C	++	++	++	++	+++	++	+	+++	++	-	-++
	NPT_2	A	++	++	++	++	+++	++	+	+++	++	-	-++
		B	++	++	++	++	+++	++	+	+++	++	-	+++
		C	++	++	++	++	+++	++	+	+++	++	-	+++
		$\frac{-}{total}$	$\frac{1}{18}$	$\frac{0}{18}$	$\frac{0}{18}$	$\frac{0}{18}$	$\frac{0}{27}$	$\frac{0}{18}$	$\frac{2}{9}$	$\frac{0}{27}$	$\frac{0}{18}$	$\frac{7}{9}$	$\frac{6}{27}$
$-Mg^{2+}$	NP	A	++	++	++	++	+++	++	-	+++	-+	-	+++
		B	++	++	++	++	+++	++	+	+++	+-	-	-++
		C	++	++	++	++	+++	++	-	++-	+-	-	+++
	NPT_1	A	++	++	++	++	+++	+-	-	+++	-+	-	+++
		B	++	++	++	++	+++	++	-	+++	++	-	-++
		C	++	++	++	++	+++	++	-	+++	++	-	+++
	NPT_2	A	++	++	++	++	+++	++	+	+++	++	-	+++
		B	++	++	++	++	+++	++	-	+++	++	-	+++
		C	++	++	++	++	+++	++	+	+++	++	-	+++
		$\frac{-}{total}$	$\frac{0}{18}$	$\frac{0}{18}$	$\frac{0}{18}$	$\frac{0}{18}$	$\frac{0}{27}$	$\frac{1}{18}$	$\frac{6}{9}$	$\frac{1}{27}$	$\frac{4}{18}$	$\frac{9}{9}$	$\frac{2}{27}$

2.4.3.4 Effect of Mg^{2+}removal on movement of insert 2

During the simulation insert 2 does not retain its interaction as predicted by the original homology model prior to molecular dynamics. Considerable movement is observed instead. Furthermore, there are some noticeable differences between the metal and non-metal simulations. In general when Mg^{2+} is removed insert 2 moves towards the trimer interface between two adjacent monomers (Fig. 2.21, Fig. 2.22). In the NP $-Mg^{2+}$ simulation insert 2 on all three monomers adopts a conformation close to the monomer interfaces of the trimer. In two chains, these inserts show reasonable agreement in that they appear to occupy the interface between two monomers. In the third chain, the insert moves to the centre of the trimer instead. In the NP $+Mg^{2+}$ simulation, this is less pronounced with some movement towards the interface between neighbouring monomers only observed for one chain. The other two remain solvent exposed. In the NPT simulations, there is movement by all three inserts

to the monomer-monomer interface with Mg^{2+}, although less pronounced than for the NP simulation. In the -Mg^{2+} simulations, the inserts again remain more solvent exposed. These results suggest insert 2 may also be involved in maintaining the trimer and may be part of the structural metal dependency. Furthermore, removing the active site metals compromises global structural integrity of *P. falciparum* arginase.

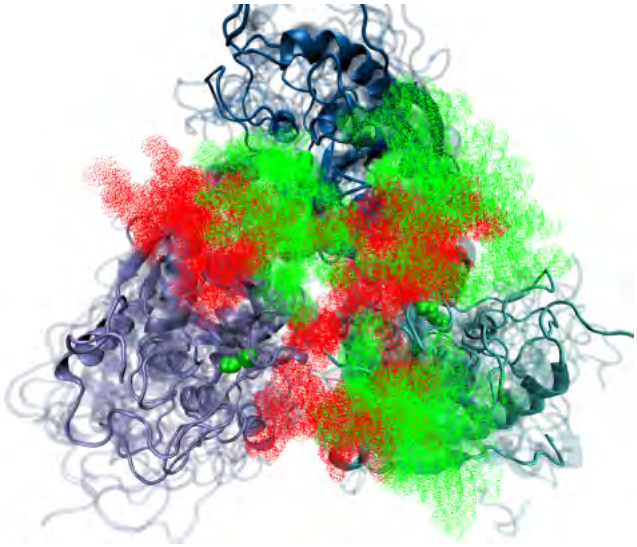


Figure 2.21: Movement of insert 2 - top view. Trimer states at the end of 20 ns (NP) and 50 ns (NPT_1 and NPT_2) are overlayed. Insert 2 is shown as clouds, Mg^{2+}(red) and non-Mg^{2+}(green). In general when Mg^{2+} is removed insert 2 moves towards the trimer interface between two adjacent monomers.

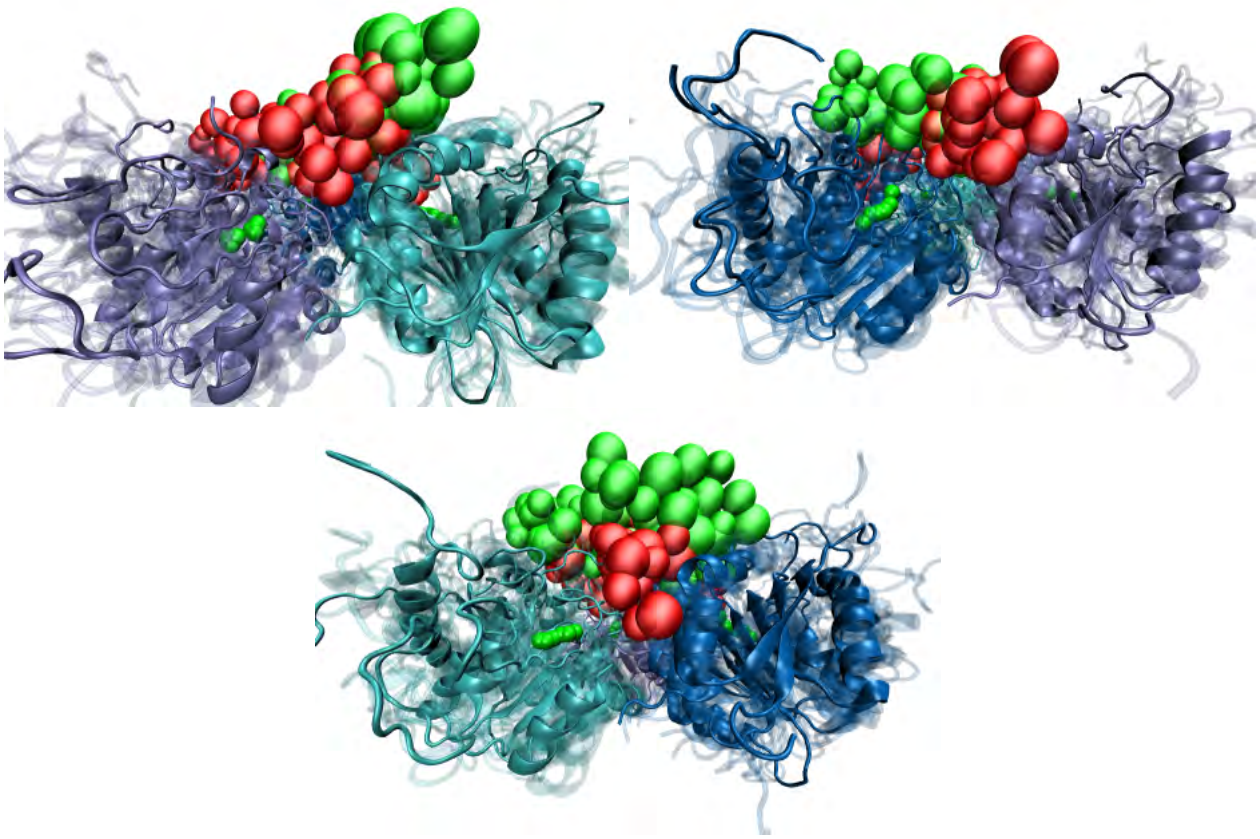


Figure 2.22: Movement of insert 2 - side on views. Trimer states at the end of 20 ns (NP) and 50 ns (NPT_1 and NPT_2) are overlayed. Insert 2 is shown as one bead per residue, +Mg^{2+}(red) and -Mg^{2+}(green). In general when $Mg^{2+}$is removed insert 2 moves towards the trimer interface between two adjacent monomers.

2.4.3.5 Integrity of inter-monomer salt-bridges

In all arginases studied to date there is a conserved inter-monomer salt-bridge represented in *P. falciparum* by Arg 346$_a$::Glu 347$_b$ (Fig. 2.9). These residues align unambiguously with their template cognates: Arg 255/274/249$_a$::Glu 256/275/250$_b$ (rat, human and bacterial templates, respectively). The salt-bridge forms reliably during modeling. Considering the established importance of this interaction, its integrity was monitored during modeling and simulation.

In the sampling runs Arg 346$_a$::Glu 347$_b$ was generally stable and intact for both the +Mg^{2+} and -Mg^{2+} case. One inter-monomer bridge did break in the Mg^{2+} case of the NP ensemble (Fig. 2.23). In the NPT ensembles the interactions mostly remained intact in both cases. One interaction was observed to break for about 6 ns in the NPT_1 -Mg^{2+} ensemble. However, there is also an increase in the average standard deviation of the salt-bridge distance in the -Mg^{2+} system in both the NPT_1 and NPT_2 simulations (Fig. 2.24 and 2.25). This suggests that Arg 346$_a$::Glu 347$_b$ is susceptible to removal of Mg^{2+}, even though the interaction remained intact.

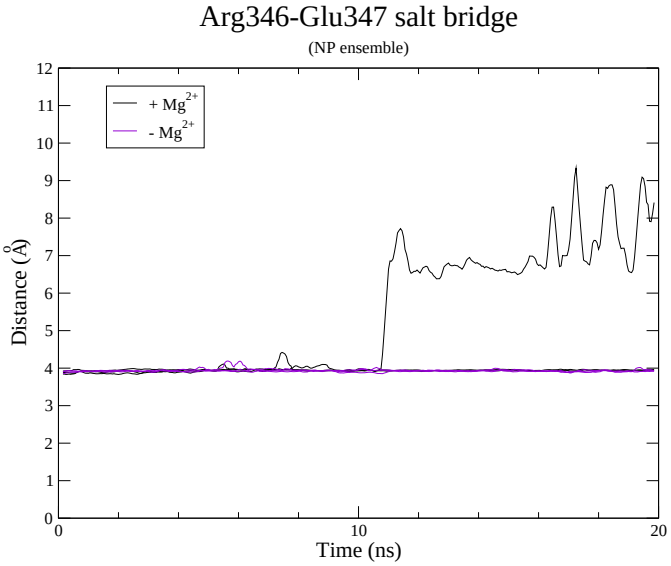


Figure 2.23: Effect of removing Mg^{2+} on the Arg 346$_a$::Glu 347$_b$ salt-bridge in the NP ensemble. One interaction is observed to break during the simulation with Mg^{2+} included. The plots were averaged using a sliding window of 500 frames (250 fs).

As described above, visual inspection of the homology models suggested a further interaction at Glu 295$_a$::Arg 404$_b$. While not fully formed in the homology models, the salt-bridge distance did adopt standard values (±4 Å) during minimisation and heating of the systems. The integrity of this interaction was found to be more susceptible to removal of Mg^{2+} than Arg 346$_a$::Glu 347$_b$. In the -Mg^{2+} systems, the salt-bridge was broken in half the possible cases over both the NP and NPT ensembles. In the NP ensemble the interaction is broken between two of the monomer pairs (between chain A and B, and chain B and C) by the end of 20 ns. The third interaction (chain C and A) was transiently broken (Fig. 2.26). In the NPT_1 ensemble two salt bridges (chain A and B, chain B and C) were broken by the end of 50 ns of simulation (Fig. 2.27), while one interaction (chain A and B) was broken by the end of 50 ns in the NPT_2 ensemble (Fig. 2.28). Since the side-chains would be freer to move in the broken interactions instead of being locked in by the salt-bridge interaction, this causes

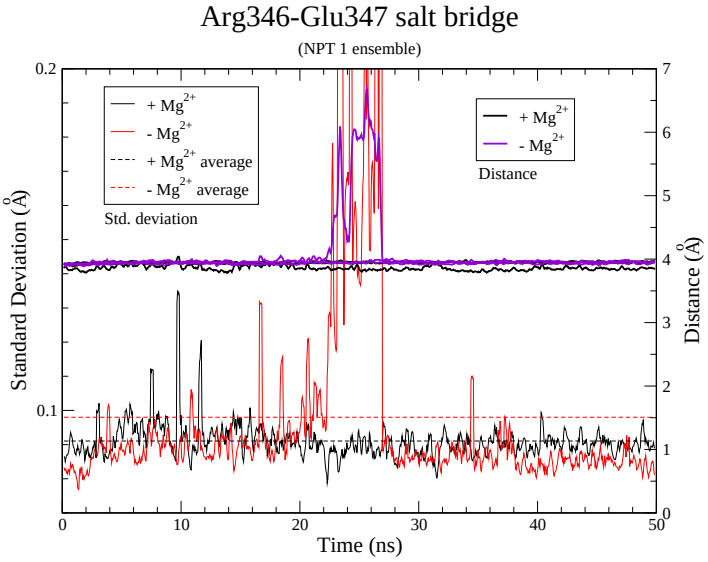


Figure 2.24: Effect of removing Mg^{2+} on the Arg 346_a::Glu 347_b salt-bridge in the NPT_1 ensemble. One interaction is observed to break briefly during the simulation without Mg^{2+}. The distances for all three salt bridge distances are overlayed: thick black (with Mg^{2+}) and thick purple (without Mg^{2+}) lines. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

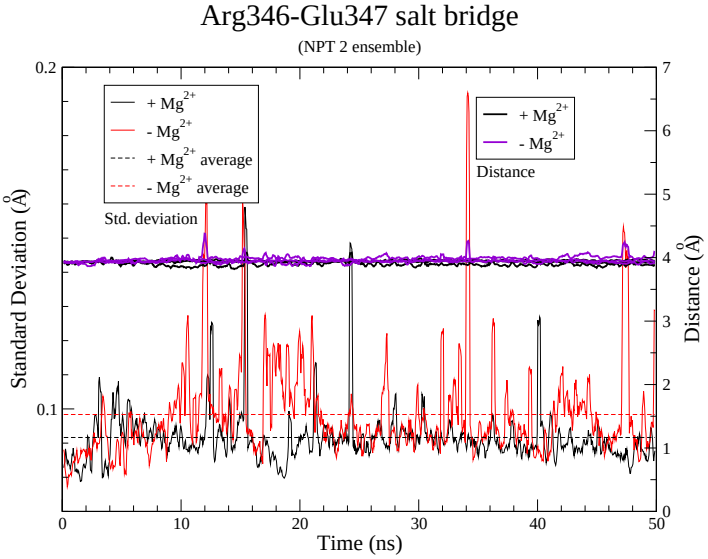


Figure 2.25: Effect of removing Mg^{2+} on the Arg 346_a::Glu 347_b salt-bridge in the NPT_2 ensemble. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

a marked increase in the standard deviation of the Glu 295_a::Arg 404_b distances in the -Mg^{2+} systems (results not shown).

In the NP ensemble, the temperature of the system was initially 310 K, but was not kept constant. During the sampling run the temperature gradually increased to approximately 332 K (Fig. 2.29). It is usual to apply some means to keep the temperature constant (isothermal ensemble), however, due to a mistake in the input files, this was omitted. Nonetheless, the results from the NP ensemble are still interesting and somewhat serendipitous. By sampling at higher temperatures, it is possible to overcome barriers in the energy landscape faster than at lower temperatures. The increase in temperature apparently accelerates the effects of removing Mg^{2+}. In the NPT simulations, only one Glu 295_a::Arg 404_b interaction is broken after 20 ns, whereas in the NP ensemble with increasing temperature all three have been broken before 20 ns. The effect of the increasing temperature is also reflected in the RMSD, which is more pronounced and more rapid in the NP ensemble. The increasing temperature may be detrimental, however, as reflected by the breaking one of the Arg 346_a::Glu 347_b

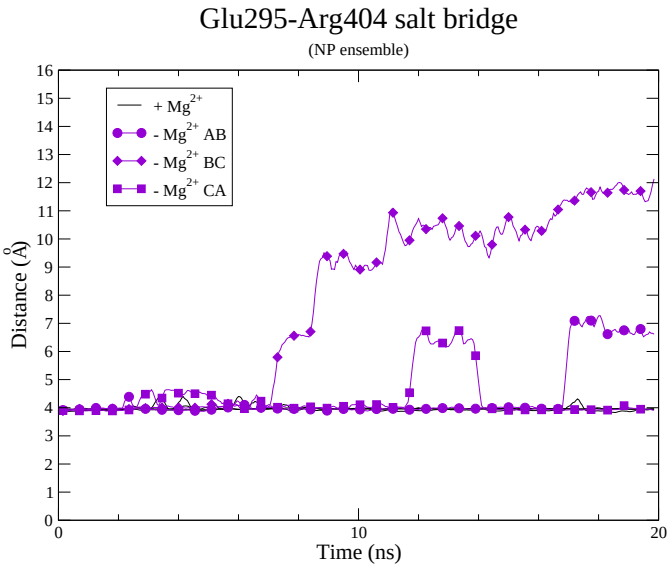


Figure 2.26: Effect of removing Mg^{2+} on the Arg 295_a::Glu 404_b salt-bridge in the NP ensemble. Without Mg^{2+} two interactions are broken at the end of the simulation, while the other is transiently broken. The three arginase monomers are arbitrarily designated A, B and C, with ◯, ◊ and □ used to indicate the three salt bridges. The plots were averaged using a sliding window of 500 frames (250 fs).

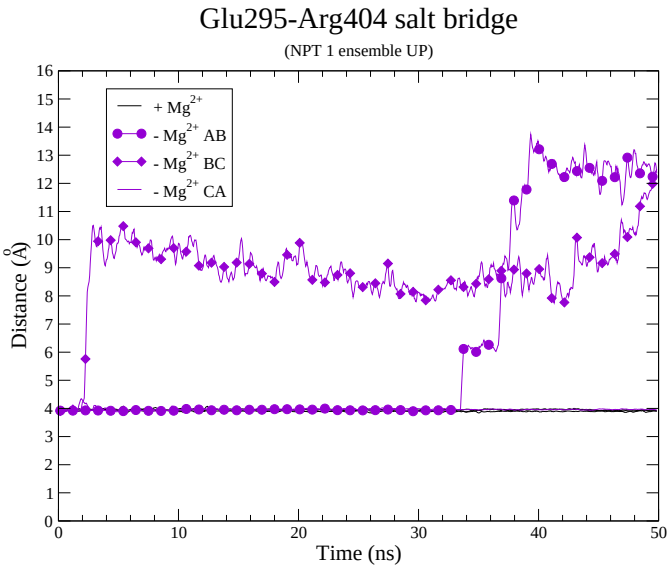


Figure 2.27: Effect of removing Mg^{2+} on the Arg 295_b::Glu 404_b salt-bridge in the NPT_1 ensemble. Without Mg^{2+} two interactions are broken at the end of the simulation. The three arginase monomers are arbitrarily designated A, B and C, with ◯ and ◊ used to indicate the two broken salt bridges. The plots were averaged using a sliding window of 500 frames (250 fs).

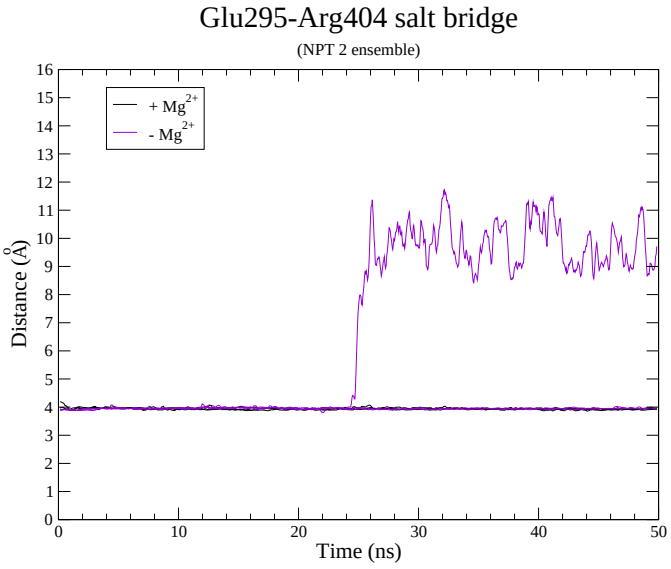


Figure 2.28: Effect of removing Mg^{2+} on the Arg 295_a::Glu 404_b salt-bridge in the NPT_2 ensemble. Without Mg^{2+} two interactions are broken at the end of the simulation. The three arginase monomers are arbitrarily designated A, B and C. The plots were averaged using a sliding window of 500 frames (250 fs).

interactions in the NP ensemble with Mg^{2+}. For this reason, all subsequent simulations were carried out in the NPT ensemble.

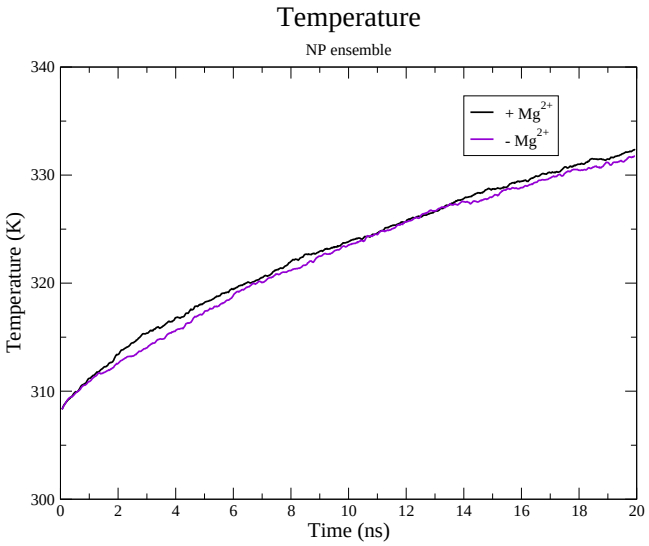


Figure 2.29: Temperature increase in the NP ensemble. When no pressure control was implemented the temperature of the system increased by the same amount in both systems with and without Mg^{2+}. The plots were averaged using a sliding window of 500 frames (250 fs).

Müller *et al.* (2005) demonstrated that mutants that abolish metal binding also disturb the trimer and that mutating the conserved interaction between Arg 346$_a$::Glu 347$_b$ results in inactive monomers. As a consequence of the simulations with and without Mg^{2+} it was decided to test the possibility of Glu 295$_a$::Arg 404$_b$ interacting experimentally. The experimental work was performed by I. B. Müller of the Bernhard-Nocht institute of Tropical Medicine, Hamburg, Germany. It was reasoned, that if Glu 295$_a$::Arg 404$_b$ is involved in this dependency, enzyme activity and trimer formation should also be susceptible to mutation. *P. falciparum* arginase was found to be more susceptible to mutations introduced at Glu 295 than Arg 404 (Table 2.4). Mutating Glu 295 to Ala or Arg considerably reduces enzyme activity by 96% and 73% respectively. Mutating Arg 404 to Ala has a less dramatic result (54% less active) but is still substantial. The double mutation of Glu 295 Ala/Arg 404 Ala also essentially inactivates the enzyme (95% less active). However, the single mutations of Glu 295 Arg and Arg 404 Ala leads to altered K_m values of 146 mM and 45 mM for L-arginine, respectively, which is up to 11-fold higher compared to the wild type arginase. The elution profile of all mutants analysed by gel filtration revealed monomeric forms, except for Glu 295 Ala, which is partially trimeric (Figure 2.30). In contrast, trimer formation is more susceptible to mutation of Arg 404, compared to enzyme activity.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 2.4: Effect of Glu 295$_a$::Arg 404$_b$ salt-bridge mutations (E295A, E295R, R404A and E295A/R404A) on arginase activity; standard error of the mean (SEM). ND: not detectable.

Mutant	V_{max} ($\mu mol.min^{-1}.mg^{-1}$)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m, ($mM^{-1}.s^{-1}$)
Wild type	31	13	24.8 (100%)	1.9 (100%)
Glu 295 Ala	1.3 ± 0.3	ND	1.0 ± 0.2 (4%)	ND
Glu 295 Arg	8.4 ± 0.9	146 ± 6	6.7 ± 0.7 (27%)	0.03 (1.6%)
Arg 404 Ala	14.3 ± 0.9	45 ± 3	11.4 ± 0.7 (46%)	0.25 (13%)
Glu 295 Ala/ Arg 404 Ala	1.6 ± 0.1	ND	1.3 ± 0.1 (5%)	ND

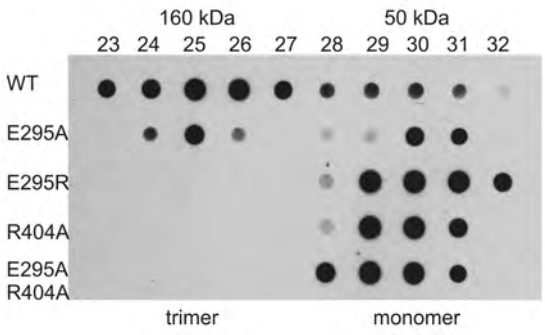


Figure 2.30: Effect of Glu 295$_a$::Arg 404$_b$ salt-bridge mutations (E295A, E295R, R404A and E295A/R404A) on trimer formation.

Thus, the existence of the inter-monomer salt-bridge at Glu 295$_a$::Arg 404$_b$ was confirmed by site-directed mutagenesis of the recombinant enzyme. All mutants tested promoted trimer dissociation, with incomplete dissociation for Glu 295 Ala. This contrasts with Glu 295 Arg which gave complete dissociation. Mutating Glu 295 to Arg is expected to be more drastic compared to Ala, because this would potentially introduce a positive charge in the vicinity of the Glu 295$_a$::Arg 404$_b$ interaction, and thus be more likely to disturb the trimer. Interestingly, this mutation leads to active but less efficient monomers with an increased K_m value of 146 mM for L-arginine, indicating altered substrate binding. In contrast, mutating Glu 295 to Ala reduced the activity to 4% of the wild-type enzyme but with its trimeric conformation partially retained. The K_m value for the Glu 295 Ala mutant was not measurable because it was not saturated up to 200 mM arginine. Mutating Arg 404 to Ala abolished trimer formation. However, this mutant enzyme shows 46% activity (as *kcat*) and 13% efficiency (as $kcat.K_m^{-1}$) and its K_m value is approximately three-fold increased compared to the wild-type enzyme. This result is similar to the previously reported behaviour of the rat liver arginase I Arg 308 mutants, which, as monomers, still had a residual activity of 41% and an efficiency in the range 13–17% (Lavulo *et al.*, 2001). Size-exclusion chromatography

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

therefore suggests that certain mutations abolish trimerisation but the enzymatic data suggests that trimerisation is not absolutely necessary for activity. However, the possibility that a weakened trimer can form under enzyme assay conditions cannot be excluded. Such a possibility is suggested by rat arginase, where the Arg 308 Lys mutant is apparently active as a monomer, but nonetheless crystallises as a trimer (Lavulo *et al.*, 2001). Although it has been demonstrated that disturbing the oligomer via the conserved Arg 346_a::Glu 347_b interaction largely inactivates the enzyme, it still has 10% residual activity (Müller *et al.*, 2005). The results of the Arg 404 Ala mutation indicates that it is possible to produce active monomers and furthermore, that certain mutations can partially compensate for induced structural instability of monomerisation by long range allosteric effects. Although there is a dependency between trimer formation and enzyme activity, these results indicate that it is not complete. This incompleteness was suggested by previous results where mutating His 193 in the active site also results in an inactive trimer (Müller *et al.*, 2005) as was also found for the Glu 295 Ala mutation in the present study. Nonetheless, formation of the predicted salt-bridge is apparently necessary for trimer formation.

The effect of mutating the Glu 295_a::Arg 404_b interaction was followed up by molecular dynamics of the mutants. During simulation, the Arg 346_a::Glu 347_b salt-bridge remains largely intact with the introduction of mutations into the Glu 295_a::Arg 404_b salt-bridge. The exceptions to this are the temporary breaking of one Arg 346_a::Glu 347_b interaction for about 10 ns in the Arg 404 Ala mutant (Fig. 2.34) and for about 5 ns in the Glu 295 Arg mutant (Fig. 2.33). In all other cases the salt-bridge distance remains at 4 Å. However, the fluctuation around the 4 Å is affected by introduction of alanine at Glu 295. In both the Glu 295 Ala and Glu 295 Ala/Arg 404 Ala mutants the average standard deviation decreases compared to the wild type (Fig. 2.31, 2.32). In contrast the introduction of Glu 295 Arg results in an increased standard deviation (Fig. 2.33), as does the Arg 404 Ala mutant. In both cases this is largely the result of the temporary breaking of one interaction (Fig. 2.34, Fig. 2.33). The introduction of the Glu 347 Gln mutant has a profound effect on Glu 295_a::Arg 404_b. All three of the interactions are broken, with one of these already disrupted during the heating stage (Fig. 2.34).

The Arg 346_a::Glu 347_b interaction can potentially interact with Glu 295 (Fig. 2.9). This is confirmed by the mutation of Glu 295 to neutral Ala which is less likely to interact strongly with either of the charged residues of Arg 346_a::Glu 347_b. In both simulations with this mutation (Glu 295 Ala and Glu 295 Ala/Arg 404 Ala) there is a lower standard deviation for the Arg 346_a::Glu 347_b distance compared to the wild type. In contrast, Glu 295 Arg causes in increase in the standard deviation compared to the wild type. Like Glu 295, Arg 295 would also be expected to interact more strongly with Arg 346_a::Glu 347_b. It is suggested that the longer sidechain of Arg compared to Glu could potentially bring it in closer contact with Arg 346_a::Glu 347_b, thus resulting in the increased effect on that interaction. Interestingly, the mutation of Arg 404 Ala also increases the standard deviation of the Arg 346_a::Glu 347_b distance. This residue is to far to make physical contact. While this might be a chance

effect, this result is in agreement with experiment which suggests that trimer formation is more susceptible to mutation at Arg 404 than Glu 295 (Fig. 2.30). The disruption of Glu 295_a::Arg 404_b by the Glu 347 Gln mutation further strengthens the suggestion that Glu 295_a::Arg 404_b is a required inter-monomer interaction, and that disruption of the trimer by mutations at Glu 347 or metal removal are mediated by this salt-bridge.

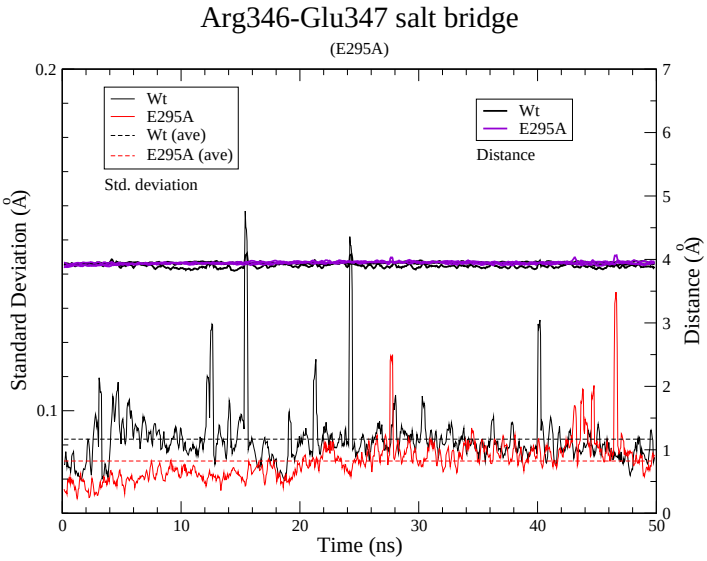


Figure 2.31: Effect of the Glu 295 Ala mutation on the Arg 346_a::Glu 347_b salt-bridge. A higher standard deviation is observed for the wild type. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

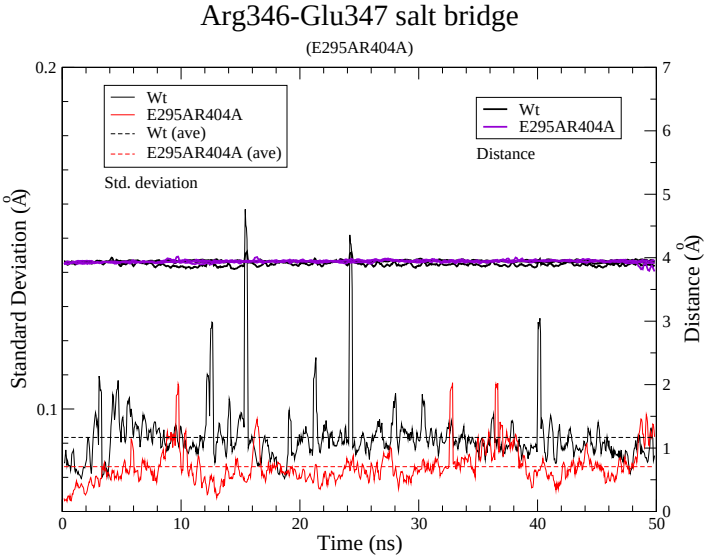


Figure 2.32: Effect of the Glu 295 Ala/Arg 404 Ala mutation on the Arg 346_a::Glu 347_b salt-bridge. A higher standard deviation is observed for the wild type. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

2.4.3.6 Co-ordination geometry of Mg^{2+}

It is expected that disturbing the interactions involved in trimer formation mediate their effects via the co-ordination of Mn^{2+} in the active site, which is required for the arginase chemistry. This is reflected by the increased equilibrium RMSD during MD, which should ultimately translate into lost co-ordination of Mn^{2+} in the active site. The loss of Mn^{2+} under such conditions, however, has yet to be observed directly. Therefore, the effect of the mutations on Mg^{2+}co-ordination during MD was analysed.

In the arginase active site, highly conserved residues are involved in a specific co-ordination pattern for the binuclear Mn^{2+} cluster. In rat I arginase the more deeply buried ion (Mn_A^{2+}) is

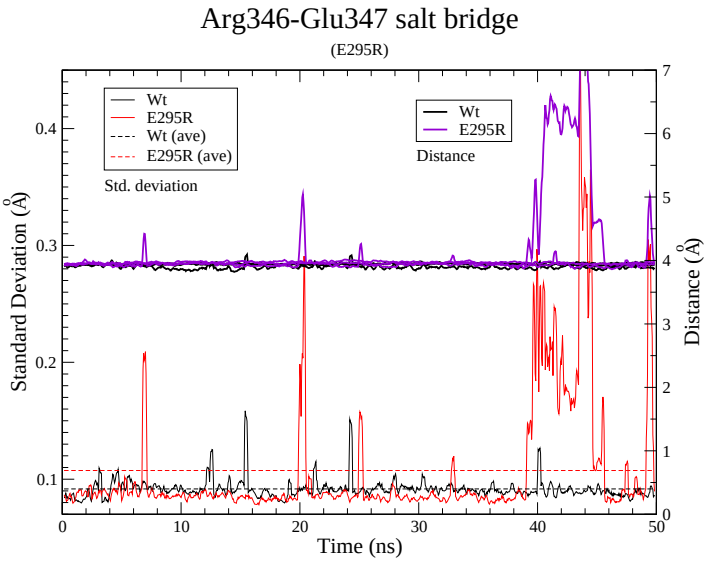


Figure 2.33: Effect of the Glu 295 Arg mutation on the Arg 346_a::Glu 347_b salt-bridge. A higher standard deviation is observed for the mutant. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

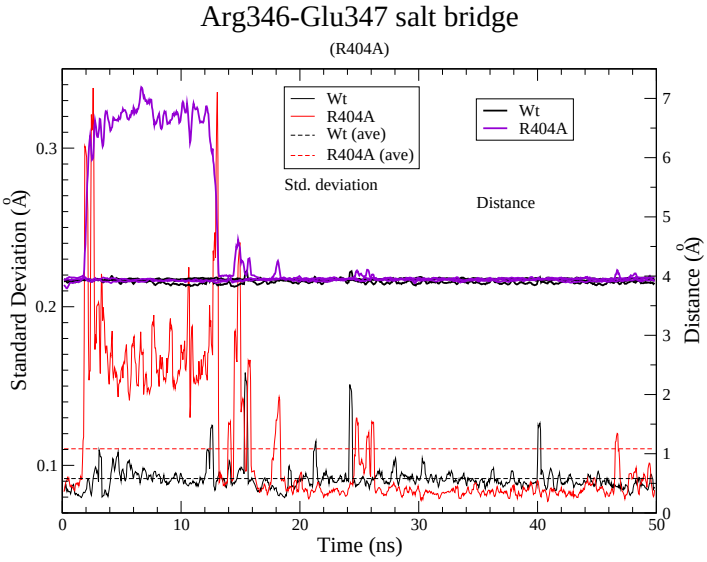


Figure 2.34: Effect of the Arg 404 Ala mutation on the Arg 346_a::Glu 347_b salt-bridge. A higher standard deviation is observed for the mutant. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

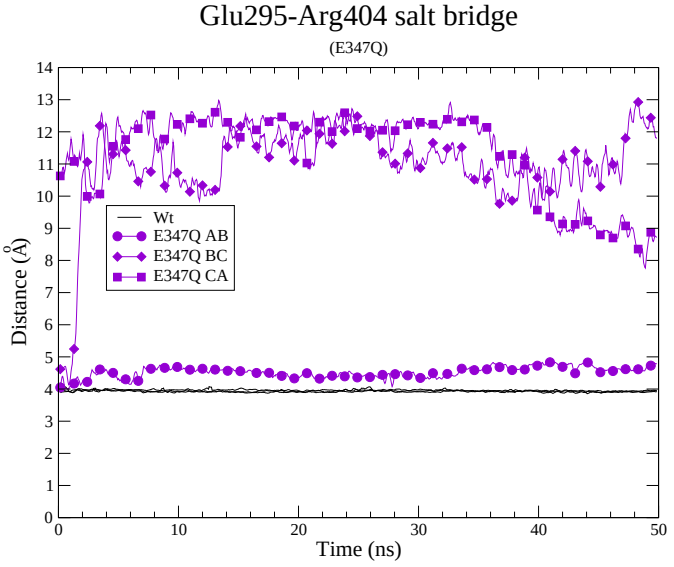


Figure 2.35: Effect of the Glu 347 Gln mutation on the Glu 295_a::Arg 404_b salt-bridge. The mutations disrupts Glu 295_a::Arg 404_b. The distance plots were averaged using a sliding window of 500 frames (250 fs). The standard deviation calculated with a sliding window of 500 and averaged for the simulations with Mg^{2+} (thin black) without Mg^{2+} (thin red). The average standard devation over the entire simulation is indicated with dashed lines.

co-ordinated by His 101, Asp 124, Asp 128, Asp 232 and the bridging solvent in a square pyramidal geometry. The respective residues in *P. falciparum* are His 193, Asp 216, Asp 220 and Asp 323. The second metal, Mn_B^{2+} is co-ordinated by His 126, Asp 124, Asp 232, Asp 234 and the bridging solvent in a distorted octahedral geometry (His 218, Asp 216, Asp 323, Asp 325 in *P. falciparum*).

During the simulations, the conformations adopted by the co-ordinating residues did not entirely conform to known crystal structures from homologues. The most notable difference is Asp 323, which is expected to form a monodentate bridging interaction between the two ions. During the simulations, it formed a bidentate bridge instead (Fig. 2.36). All other expected co-ordinating atoms were oriented close enough to interact with the ions. The only other missing interaction was that of the bridging OH^- as no attempt was made to introduce the bridging solvent molecule. The Mg^{2+}-Mg^{2+} distance was also about 0.6 Å greater than the known Mn^{2+}-Mn^{2+} distance. The larger distance is partly due to the inability of the software to recognise co-ordination chemistry natively as well as the larger van der Waals radius of Mg^{2+} compared to Mn^{2+}. The Mg^{2+} ions nonetheless remained in the active site during the simulations and restricted the movement of the interacting ligands.

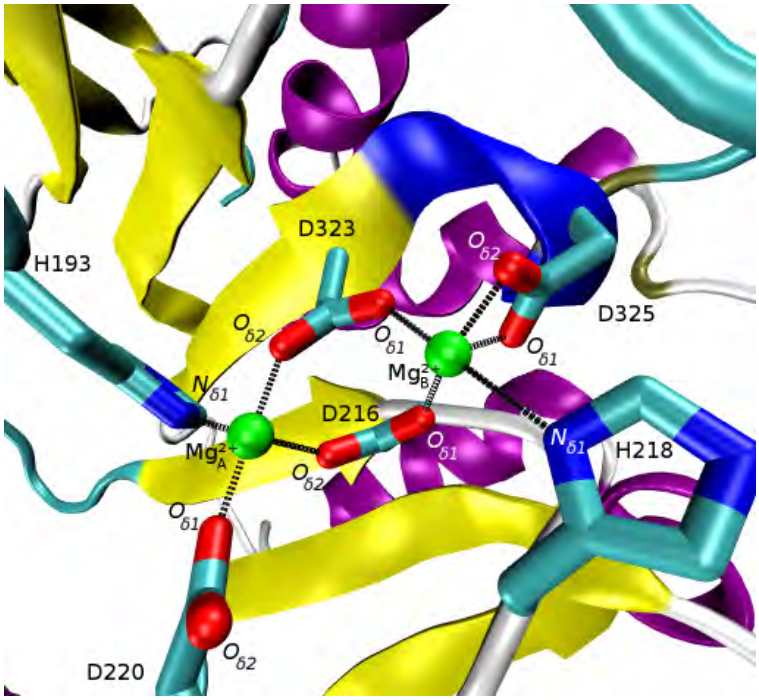


Figure 2.36: General co-ordination pattern of Mg^{2+} in the NPT_2 ensemble in chain A after 40 ns. The bridging OH^- is absent. Mg_A^{2+} is co-ordinated in an incomplete trigonal bipyramidal geomentry, while Mg_B^{2+} is co-ordinated in an incomplete octahedral geometry. The missing ligand is an OH^- which could not be included for modeling due to force-field restrictions. The exact geometry also differ slightly compared to experimental structures. The presence of Mg^{2+} still restricts ligand movement, however.

The stability of the active site was observed by monitoring the inter-atom distances between the interacting residues and Mg^{2+}. Four interactions occur between Mg_A^{2+} and His $193N_{\delta 1}$, Asp $220O_{\delta 1/\delta 2}$, Asp $216O_{\delta 1/\delta 2}$ and Asp $323O_{\delta 1/\delta 2}$. Five interactions occur between Mg_B^{2+} and His $218N_{\delta 1}$, Asp $216O_{\delta 1/\delta 2}$ Asp $323O_{\delta 1/\delta 2}$, Asp $325O_{\delta 1}$ and Asp $325O_{\delta 2}$. For the acidic residues Asp 323, Asp 216 and Asp 220 generally only one of the chemically equivalent carboxyl oxygens ($O_{\delta 1}$ or $O_{\delta 2}$) interacts with an Mg^{2+} ion. Although these oxygens may potentially exchange places in interacting with Mg^{2+}, this was never observed during any of the solvated simulations. The interacting carboxyl oxygens generally maintained a constant distance of ± 1.8 Å with the Mg^{2+} ion, while the non-interacting oxygens displayed

greater variability. During the NPT_2 simulation an apparent partial exchange of position was observed for Asp 220 with Mg_A^{2+} in chain C where both $O_{\delta 1}$ and $O_{\delta 2}$ distances reach ± 1.9 Å. The active sites of the NPT simulations were generally stable during 50 ns simulation, apart from the altered conformation of Asp 220 described above, as well as a slight increase for the His 193 $N_{\delta 1}$-Mg_A^{2+}interaction, also in chain C. In the NP simulation, there was more instability. The distances between Mg_A^{2+} and both $O_{\delta 1}$ and $O_{\delta 2}$ of Asp 216 increased to about ± 5.8 Å from ± 1.8 Å by the end of 20 ns in chain C (Fig. 2.37). Furthermore, the distance between His 218 $N_{\delta 1}$ and Mg_B^{2+} increased to ± 7 Å (Fig. 2.38). The Asp 216 and His 218 interactions with their respective Mg^{2+} were thus essentially broken. The greater instability during the NP simulation is thought to be due to the increase in temperature that occurred (explained below). It is expected that the interactions with Mg^{2+} in the active site should remain stable, since the bi-nuclear cluster is required for enzyme activity.

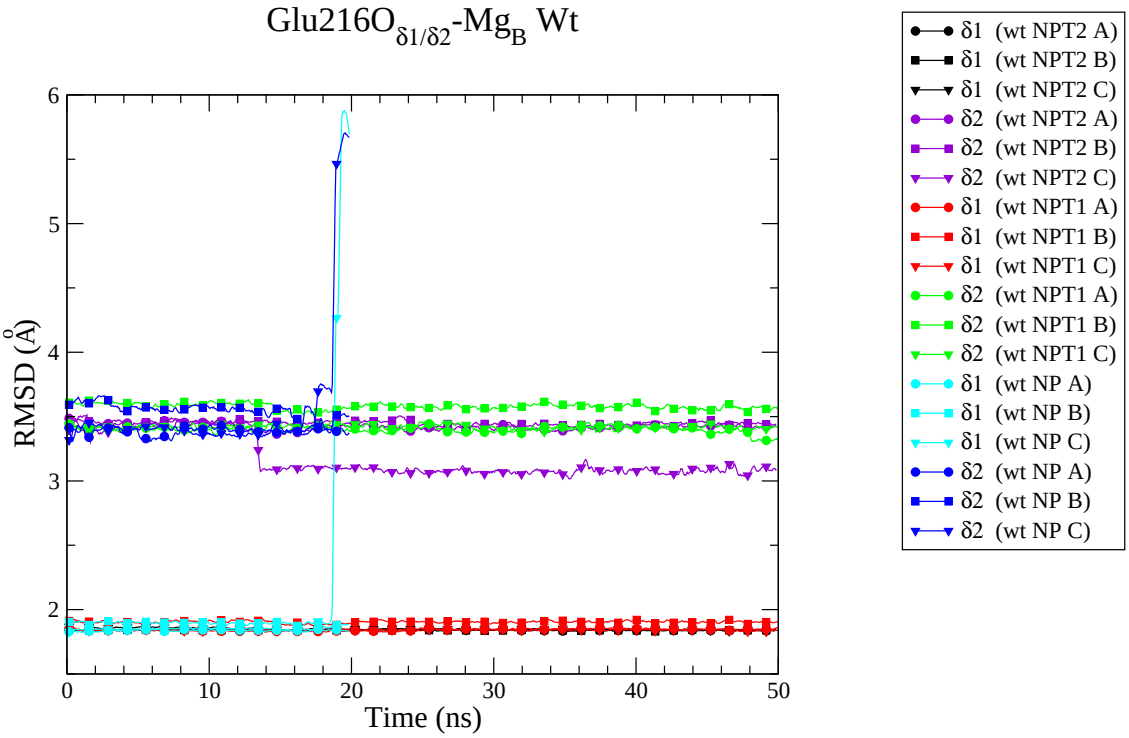


Figure 2.37: Distance between Asp 216 $O_{\delta 1/\delta 2}$ and Mg_B^{2+} in wild type arginase in the NP, NPT_1 and NPT_2 simulations. Both carboxyl oxygens ($\delta 1/\delta 2$) are included. Typically only one carboxyl O interacts with Mg^{2+} at a distance of ± 1.8 Å, leaving the other carboxyl O to vary at ± 3.4 Å. Most interactions remained stable. One residue interaction was broken in chain C of the NP simulation. Pairs of carboxyl O are indicated for chains A (○), B (□) and C (▽).

When simulating without Mg^{2+} removed, considerable movement is observed in the co-ordinating residues in all three simulations (NP, NPT_1 and NPT_2). However, in the NP simulation there is more more movement of the co-ordinating residues with Mg^{2+} compared to the NPT simulations. This is likely to be due to the increased temperature during the simulation (Fig. 2.39).

Because the presence of Mg^{2+} was able to stabilise the co-ordinating residues by electro-

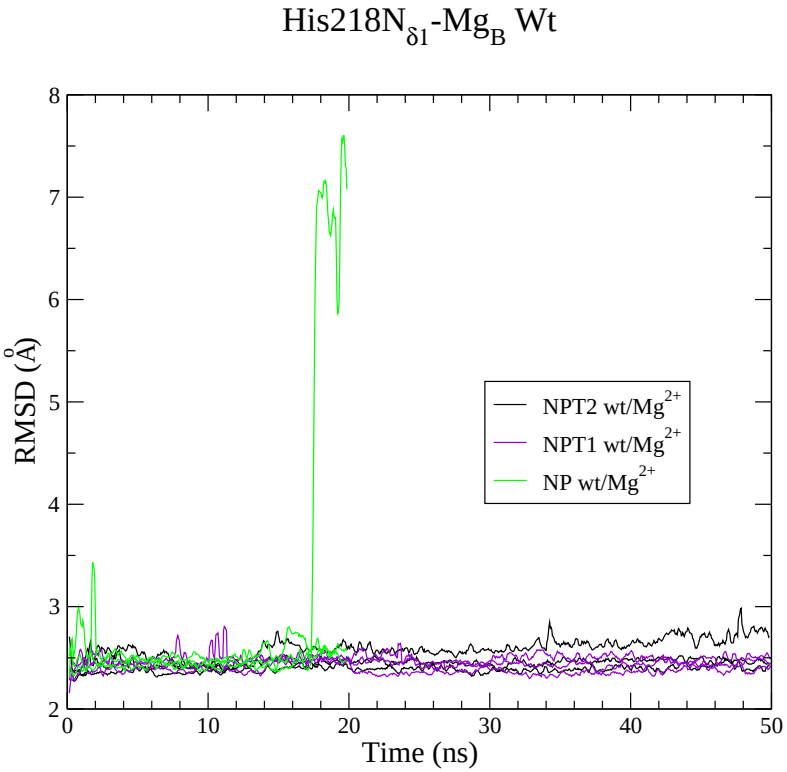
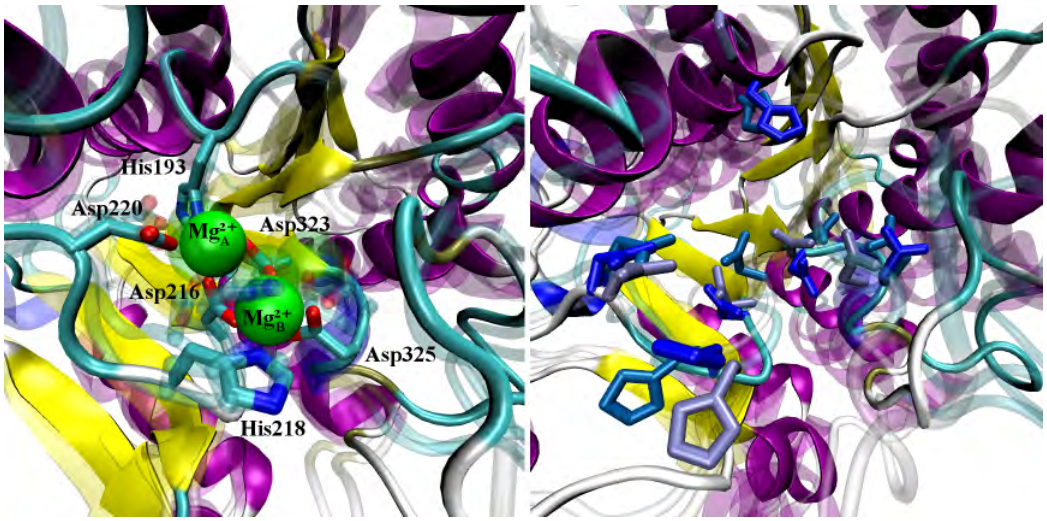


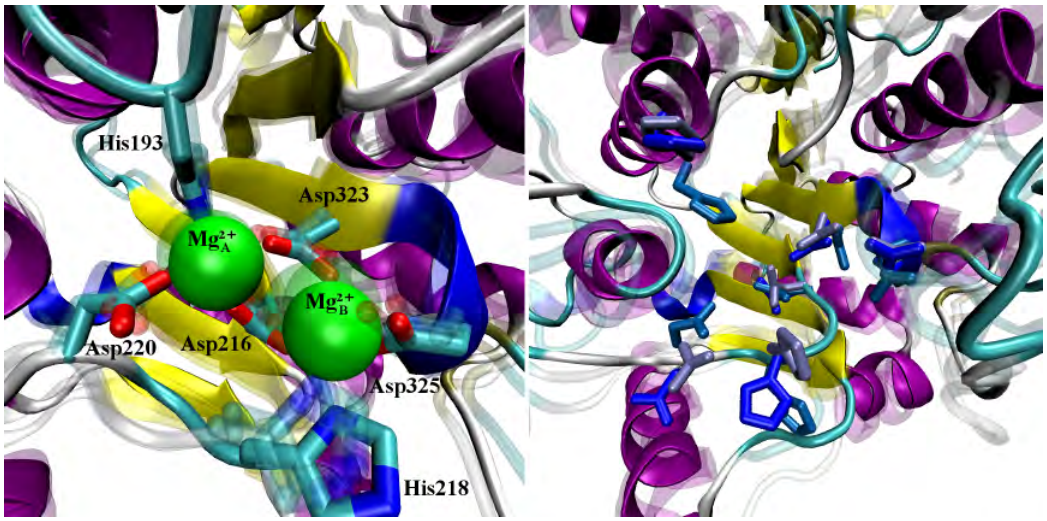
Figure 2.38: Distance between His 218$N_{\delta1}$ and Mg_B^{2+} in wild type arginase in the NP, NPT_1 and NPT_2 simulations. Most interactions remained stable. One interaction is broken in chain C of the NP simulation. The plots for chains A, B and C are overlaid.

static interactions alone, this approach appears viable for investigating the structural metal dependency. This suggests that structural metal dependency involves free movement of the metal co-ordinating residues. The extra stability of the active sites in the NPT ensembles compared to the NP ensemble, suggests that this is the more appropriate type of simulation.

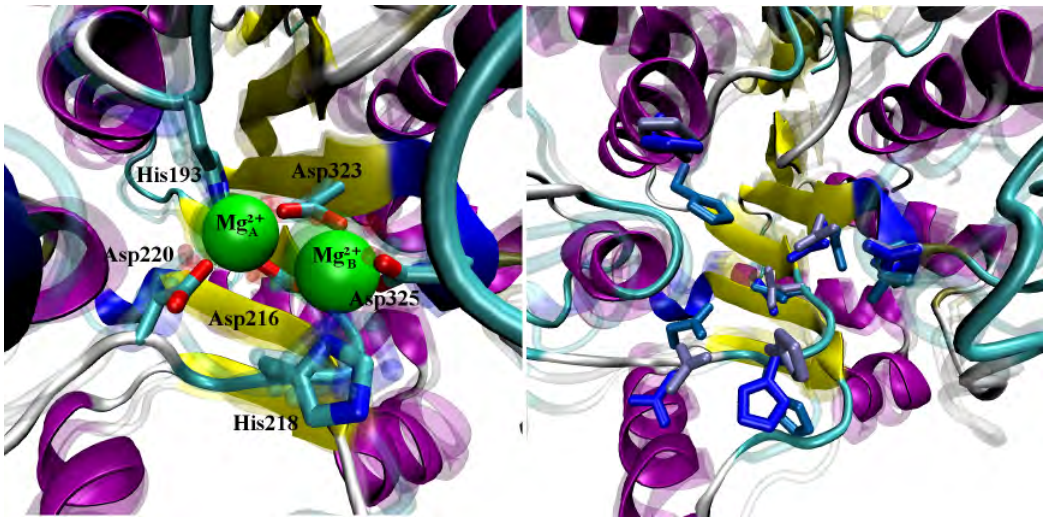
During the simulation of the various mutants, disturbances were observed in the active site compared to the wild type simulations. In three out of five mutants there were large deviations in the interaction distances compared to the standard distances observed during the wild type simulations. No interactions were lost during the simulation of Arg 404 Ala and Glu 295 Ala/Arg 404 Ala. In the Glu 295 Ala simulation the distance between Asp 216$O_{\delta1/\delta2}$ and Mg_B^{2+} increases to 5.0-5.5 Å (from ± 1.8 Å) for about 6 ns (Fig. 2.40) and the distance between His 218$N_{\delta1}$ and Mg_B^{2+}increases to 4.5-7.5 Å (from ± 1.8 Å) for about 8 ns (Fig. 2.41). In Glu 295 Arg and Glu 347 Gln mutants more drastic increases in the interaction distances were observed. In Glu 295 Arg the distance between Asp 216$O_{\delta1/\delta2}$ and Mg_B^{2+} increases to ± 4.5-6 Å for about 30 ns in chain B (Fig. 2.42). Furthermore – also in chain B – the distance between His 218$N_{\delta1}$ and Mg_B^{2+} increases from ± 5 Å to ± 8 Å during 40 ns (Fig. 2.43). The distance had already increased to ± 5 Å during the heating stage. The greatest number of disturbances are observed for Glu 347 Gln. A number of these are observed early in the production run, suggesting that mutating Glu 347 to Gln introduces a significant disturbance. Firstly, the distance between Asp 323$O_{\delta1/\delta2}$ and Mg_A^{2+} increases to about ± 3.8



(a) NP



(b) NPT_1



(c) NPT_2

Figure 2.39: Destabilisation of active site metal co-ordinate residues. All three chains within each of the $+Mg^{2+}$(left) and $-Mg^{2+}$(right) from the NP, NPT_1 and NPT_2 simulations were superimposed. Two out of the three chains are represented as transparent, except for the co-ordinating residues for the $-Mg^{2+}$simulations, which are indicated in different shades of blue. When Mg^{2+} is removed there is greater movement of the co-ordinating residues compared to when $Mg^{2+}$is present.

Å and ± 5.1 Å for each of the sidechain carboxyl oxygens for 40 ns in chain C (Fig. 2.44). Secondly, the interaction between His 218 $N_{\delta 1}$ and Mg_B^{2+} increases to ± 6-8 Å for chains B and C for about 39 ns each (Fig. 2.45). The interactions for which there were no significant deviations are included in Appendix A.

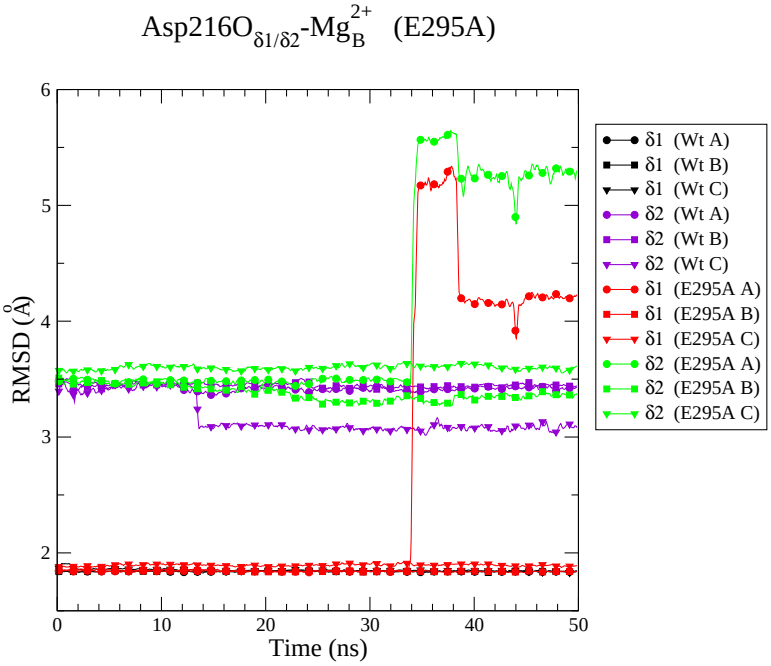


Figure 2.40: Interaction between Asp 216 O and Mg_B^{2+} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta 1/\delta 2$) are included. Pairs of carboxyl O are indicated for chains A (○), B (□) and C (▽). In the Glu 295 Ala mutant one interaction is broken (chain C) for the last six ns.

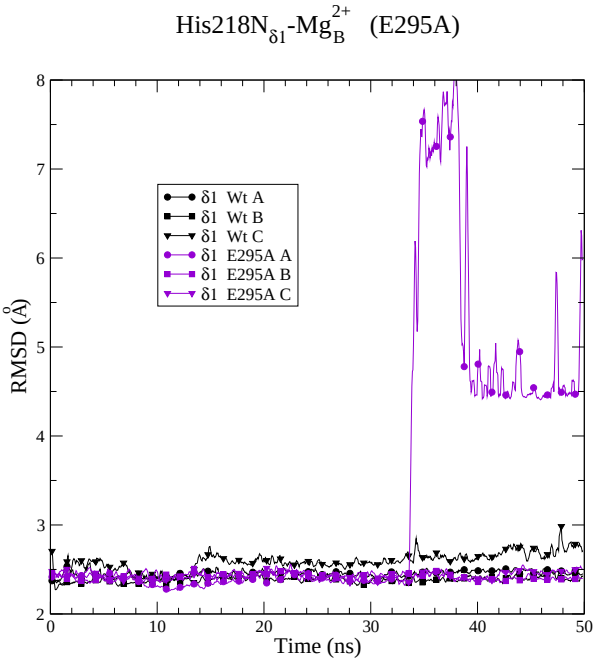


Figure 2.41: Interaction between His 218 $N_{\delta 1}$ and Mg_B^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Plots are indicated for chains A (○), B (□) and C (▽). In the Glu 295 Ala mutant one interaction is broken (chain C) for the last eight ns.

It has been demonstrated certain mutations to *Pf* Arg located away from the active site and partaking in trimer formation, compromise activity. Furthermore, mutations that abolish metal binding, or removal of the catalytic metal, compromise activity (Müller *et al.*, 2005). Therefore, it is suggested that mutations to the inter-monomer salt-bridges mediate their effects by causing the loss of metal binding. The simulations with of the *Pf* Arg mutants suggests this may be the case. During the simulation of certain mutants, there was partial

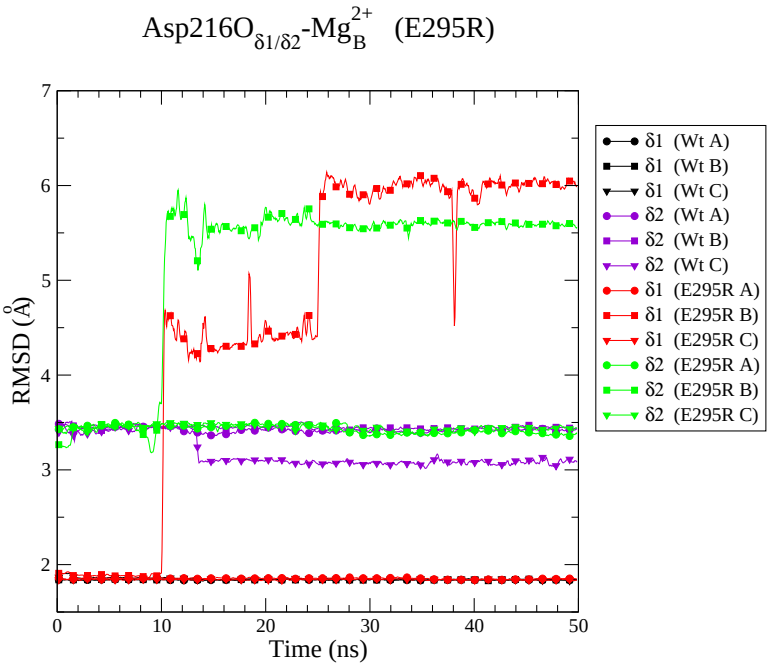


Figure 2.42: Interaction between Asp 216 O and Mg^{2+}_{B} in *pf*Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$). In the Glu295Arg mutant one interaction is broken (chain B) for the last 30 ns.

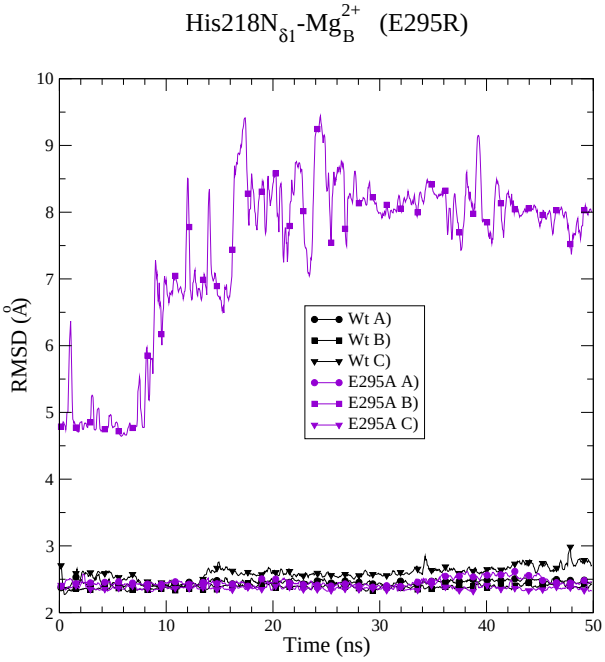


Figure 2.43: Interaction between His 218 $N_{\delta1}$ and Mg^{2+}_{B} in *pf*Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$). In the Glu 295 Arg mutant one interaction is broken (chain B) for 40 ns.

loss of the interactions expected to maintain the metals in the active site. Loss of these interactions is likely to lead to loss of metal binding and thus activity. During the simulations, the greatest disturbance to interactions with Mg^{2+} was observed for Glu 347 Gln, Glu 295 Arg and Glu 295 Ala, whereas little disturbance was observed for mutations to Arg 404 (Arg 404 Ala and Glu 295 Ala/Arg 404 Ala). This is in partial agreement with experimental results concerning these mutations. Loss of activity is observed for all these mutations, with the least reported loss for Arg 404 Ala (Table 2.4) and the greatest loss for mutations to Glu 295 or Glu 347 (Müller *et al.*, 2005). However, although the double mutant Glu 295 Ala/Arg 404 Ala results in 95% loss of activity, simulations of this mutant did not reveal any disturbance to the active site. Furthermore, although there were disturbances to the active site during the Glu 295 Ala simulation, these occurred only within one chain for a relatively brief period.

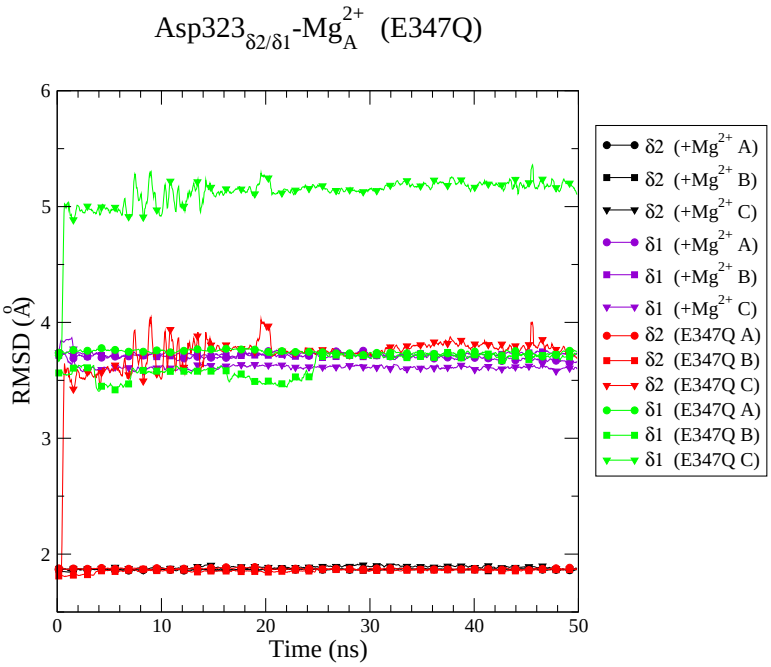


Figure 2.44: Interaction between Asp 323 O and Mg_A^{2+} in *pf*Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta 1/\delta 2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$). In the Glu347Gln mutant one interaction is broken (chain C) for the last 30 ns.

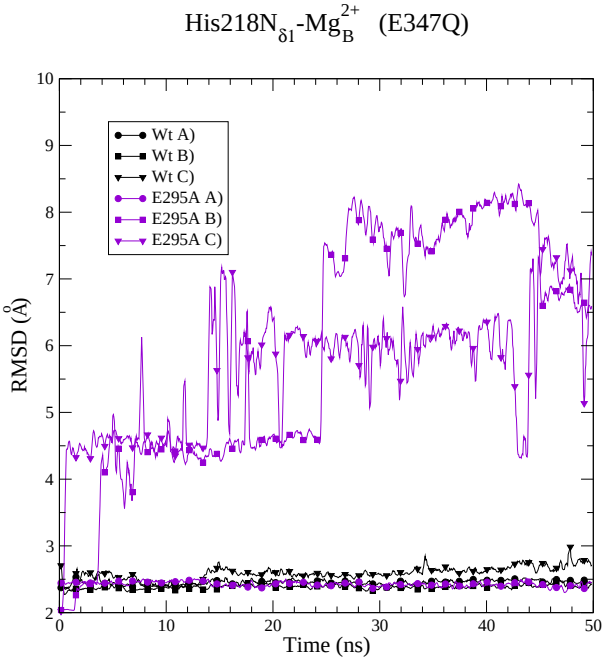


Figure 2.45: Interaction between His 218 N and Mg_B^{2+} in *pf*Arg Glu 347 Gln compared to wild type. Plots are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$). In the Glu347Gln mutant two interactions are broken (chains B, C) for about 39 ns each.

Thus, the simulations performed do not completely capture what is observed experimentally. The times simulated were relatively short (50 ns) and thus may need to be extended to obtain agreement with experiment.

2.4.3.7 Movement per residue

In order to determine whether any part of the protein reproducibly moved more upon removal of the active site metals, the RMSD per residue was analysed. This was carried out in VMD by aligning the whole protein of each frame of the sampling run to the first frame. The change in RMSD over each residue was determined for each frame and summed. This value was then averaged by the number of frames. In the NP simulation, certain portions of the protein that form part of the inter-monomer interfaces were observed to have a higher RMSD

per residue (Fig. 2.46). However, this was not reproducible in the NPT simulations.

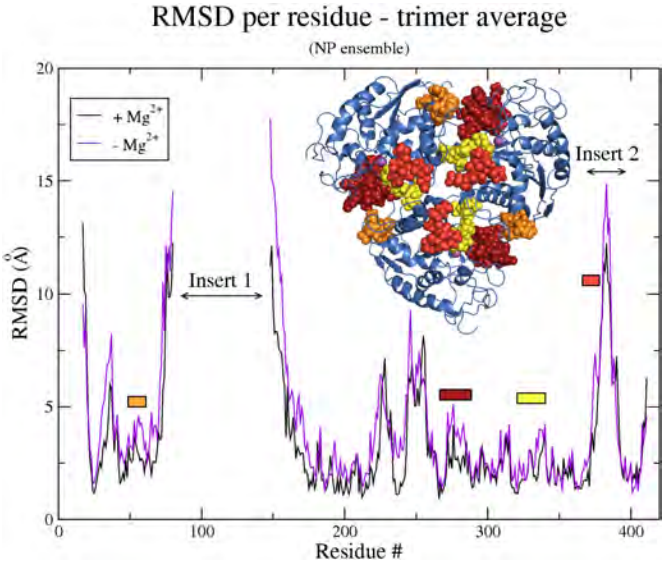


Figure 2.46: RMSD per residue for the NP simulation. When Mg^{2+} is removed certain regions are observed to move more than with Mg^{2+}. Among these regions are inter-monomer contacts indicated in the structure. Each chain was superimposed (all protein atoms) independently. The average for all three chains in each system (with and without Mg^{2+}) is shown. All frames from the simulation were used for the calculation.

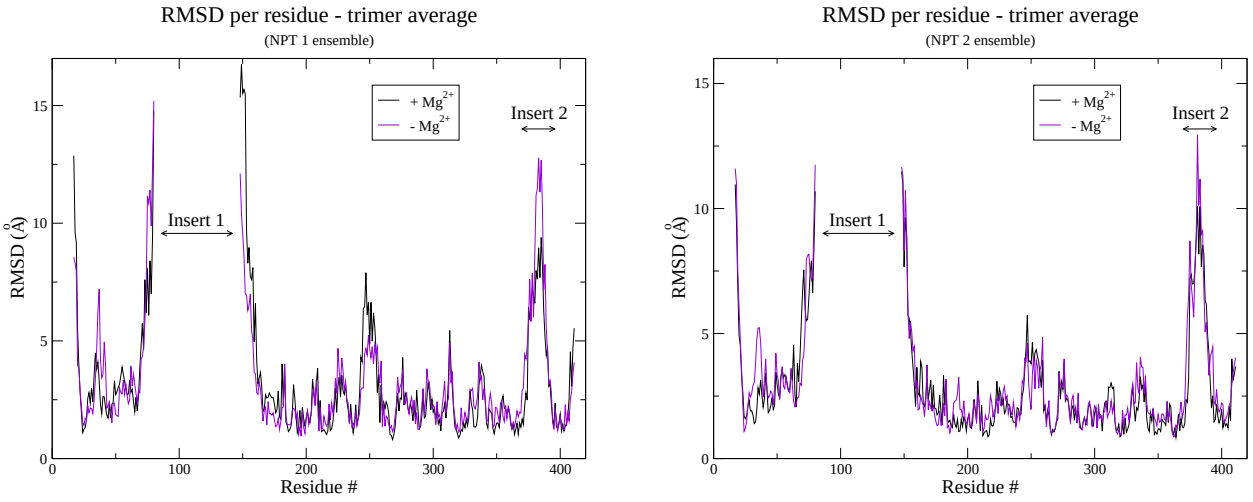


Figure 2.47: RMSD per residue for the NPT simulations. Compared to the NP simulation the same increase in movement of certain interacting regions without Mg^{2+} is not observed. Each chain was superimposed (all protein atoms) independently. The average for all three chains in each system (with and without Mg^{2+}) is shown. All frames from the simulation were used for the calculation.

The RMSD per residue results are generally mirrored when calculating the root mean square fluctuation (RMSF) instead, which is the standard deviation of the RMSD. In the NP simulation the RMSF per residue results mirror that of RMSD per residue (Fig. 2.48). RMSF of the $C_\alpha$atoms gives less noisy plots compared to RMSD per residue, and it becomes possible to observe increased movement in certain regions in the NPT simulations (Fig. 2.49). The increased movement of certain interface contacting regions observed in the NP simulation is partially repeated in the NPT simulations, however. There is increased movement for insert 2 in all three simulation, in agreement with increased movement of insert 2 towards the interface regions without Mg^{2+}.

The increase in RMSD observed in the NP ensemble but not repeated for NPT could be due to the increase in temperature that occurred in the NP simulation. Furthermore,

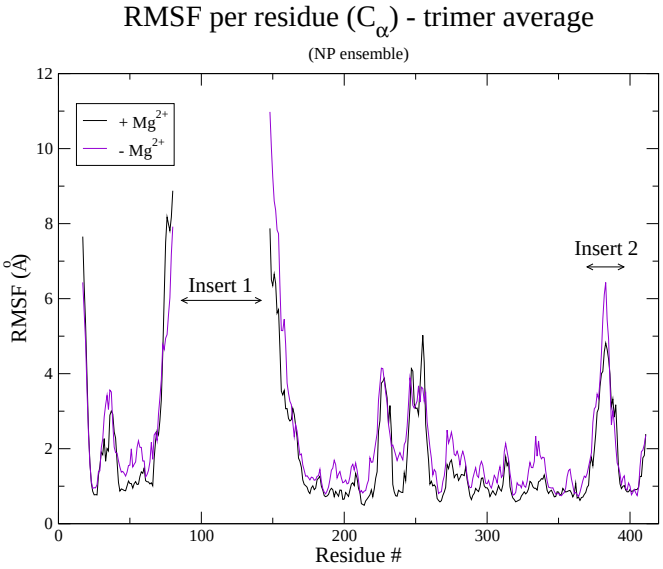


Figure 2.48: Root mean square flucation (RMSF) or standard deviation of RMSD per residue for the NP simulation (C_α atoms). This generally mirrors the RMSD per residue results. Each chain was superimposed (all protein atoms) independently. The average for all three chains in each system (with and without Mg^{2+}) is shown. Every 100th frame was used (400 frames total) for the calculation.

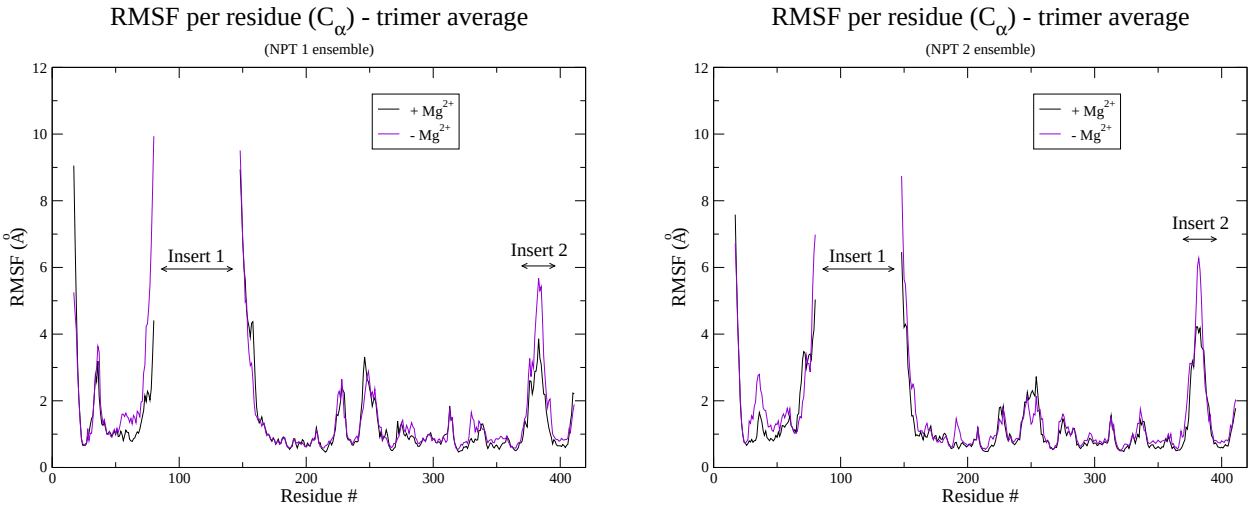


Figure 2.49: RMSF per residue (C_α atoms) for the NPT simulations. Compared to the NP simulation the same increase in movement of certain interacting regions without Mg^{2+} is not observed. Each chain was superimposed (all protein atoms) independently. The average for all three chains in each system (with and without Mg^{2+}) is shown. Every 100th frame was used (1000 frames total) for the calculation.

when measuring RMSF instead of RMSD certain regions that were observed to move more with Mg^{2+}removed in the NP ensemble could also be detected for NPT. Thus, RMSF seems to be a more sensitive measure for detecting differences between systems with and without Mg^{2+}. However, to obtain full agreement between NP and NPT it is expected that longer runs would be required. Due to time and hardware constraints this could not be attempted.

2.5 Conclusion

Increasing drug resistance to malaria highlights the need for new treatments. Lately, interfering with protein-protein interactions is being considered for potential drug discovery (Fry, 2006; Fletcher and Hamilton, 2006; Keskin *et al.*, 2007). *P. falciparum* arginase is notably different from the host equivalent in that enzyme activity and trimer formation are mutually

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

inclusive. In mammals it has been demonstrated that rat arginase I loses some activity (33–41% of *kcat*) when the trimerisation is disturbed by mutagenesis (Lavulo *et al.*, 2001). This has not been observed for human arginase I, where fully functional monomers have been obtained (Sabio *et al.*, 2001; Carvajal *et al.*, 1971, 1977). Despite the high sequence similarity (87%) between arginase I from rat and human (Di Costanzo *et al.*, 2007b), they differ in their kinetic properties. Human arginase I has a substantially lower K_m for arginine compared to rat arginase I. Furthermore, the K_d values for the inhibitors *S*-(2-boronoethyl)-L-cysteine and 2-amino-6-boronohexanoic acid are one order of magnitude less than for the rat counter part (Cavalli *et al.*, 1994; Baggio *et al.*, 1999; Kim *et al.*, 2001; Di Costanzo *et al.*, 2005). This suggests that it may be possible to inhibit *Pf*Arg via disturbing oligomerisation without affecting the human counterpart.

Arginine levels in *in vitro* cultures of *P. falciparum* are depleted by *Pf*Arg, although the relevance of arginase as a malaria drug target remains to be demonstrated (Olszewski *et al.*, 2009). Hypoargininaemia has been linked to the progression of severe malaria and may be related to the requirement of arginine for NO biosynthesis (Weinberg *et al.*, 2008). It has been speculated that low host NO benefits the parasite by causing increased expression of host intracellular adhesion molecule-1, which is used by parasitised red blood cells to adhere to the vascular endothelium and thus avoid spleen clearance. Arginase knockouts of the rodent malaria parasite, *P. berghei* (ANKA strain), are viable and show similar growth behaviour *ex vivo* and in infected mice (Olszewski *et al.*, 2009), although this has yet to be established for the human parasite.

Nonetheless, interfering with trimer formation in *Pf*Arg may serve as a parasite specific means of enzyme inhibition. To this end, *P. falciparum* arginase was modelled to further understand the structural metal dependency. Visual analysis of the model revealed a possible new inter-monomer salt-bridge between Glu 295 and Arg 404. Molecular dynamics simulations suggest that this interaction is susceptible to the removal of the active site metals. This was further corroborated by site-directed mutagenesis of the recombinantly expressed enzyme. Mutations expected to abolish the interaction promoted dissociation of the trimer and decreased enzyme activity. Simulations of *P. falciparum* arginase without the active site metals also resulted in a larger equilibrium C_α-RMSD, thus indicating that absence of the metals causes protein instability. However, the simulations were not long enough to observe dissociation or substantial protein unfolding. Longer simulations could therefore be pursued to further understand the possible linkage between active site metal binding and the Glu 295_a::Arg 404_b interaction. A further important result was that the general loss of structural integrity could be observed across different MD protocols, suggesting a real biological effect was observed and that this was not an artifact of the simulation. While some differences were observed between the NPT_1 and NPT_2 protocols, there was not sufficient time and resources to see if these could be replicated.

The results suggest that Glu 295_a::Arg 404_b interaction is necessary for trimer formation and enzyme activity of *P. falciparum* arginase, and forms part of the mechanism of

the structural metal dependency. The system used to simulate was incomplete in that it replaces Mn^{2+} with Mg^{2+} and does not completely simulate the co-ordination chemistry. Nonetheless, the results indicate this system could be used for *in silico* testing of molecules that interfere with trimer formation. Interfering with the formation of the Glu 295_a::Arg 404_b salt bridge may provide for novel inhibitors, since this interaction is not present in the host enzyme. The model system described could be used for *in silico* screening of small molecules designed to disturb the Glu 295_a::Arg 404_b and Arg 346_a::Glu 347_b interactions. Specifically, as Glu 347 Gln disturbs Glu 295_a::Arg 404_b suggests that a compound that induces similar behaviour may disrupt the enzyme.

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.

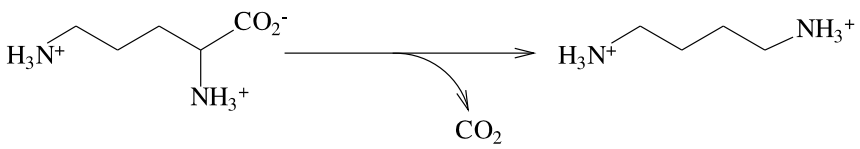


Figure 3.1: Ornithine decarboxylase reaction

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 ($\beta 12 \uparrow -\beta 13 \downarrow -\beta 15 \downarrow -\beta 16 \uparrow$).

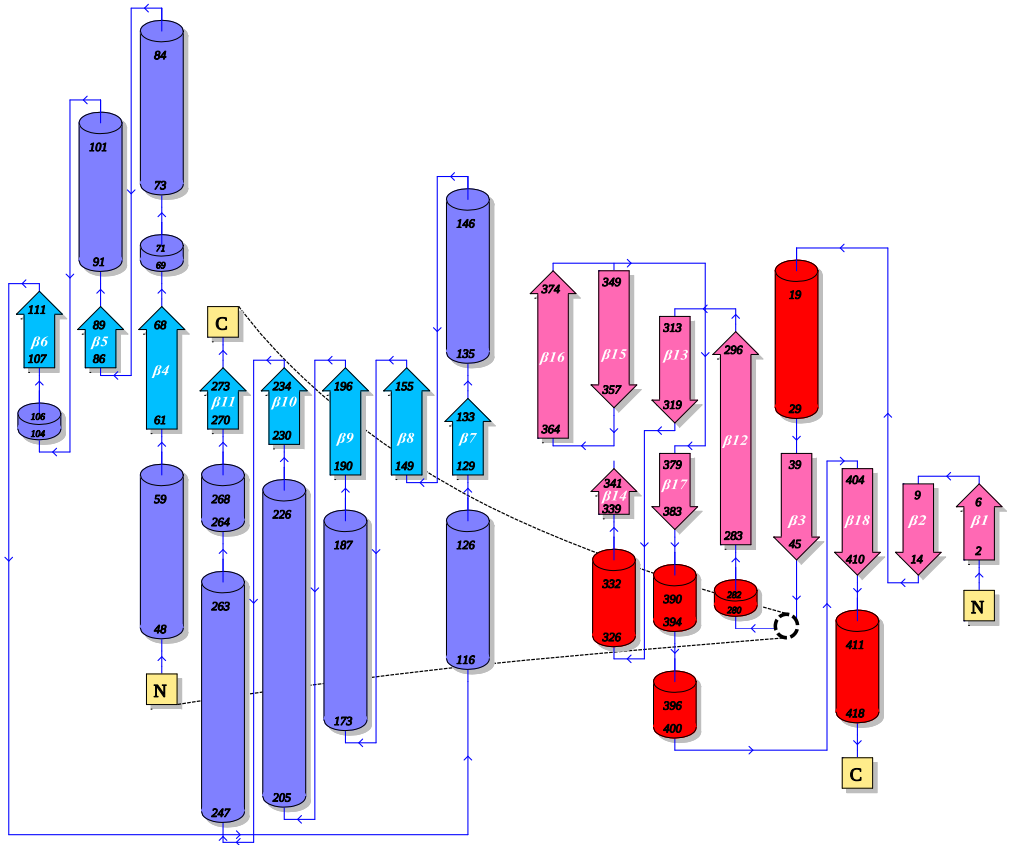


Figure 3.2: Topology of ODC TIM barrel (left) and sheet (right) domains, PDB id: 7ODC. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The insertion site of the TIM 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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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).

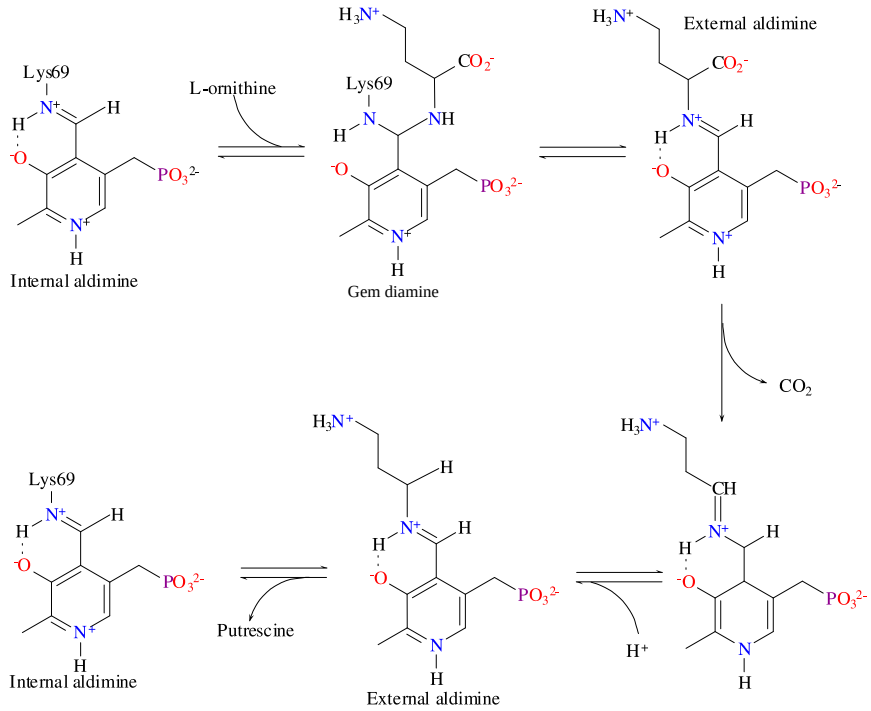


Figure 3.3: Reaction mechanism of ornithine decarboxylase, adapted from Jackson *et al.* (2003a). The incoming substrate displaces Lys 69 to form the external aldimine via a gem diamine intermediate. The reaction that follows depends on the orientation of the moieties around C_α relative to the PLP plane. 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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

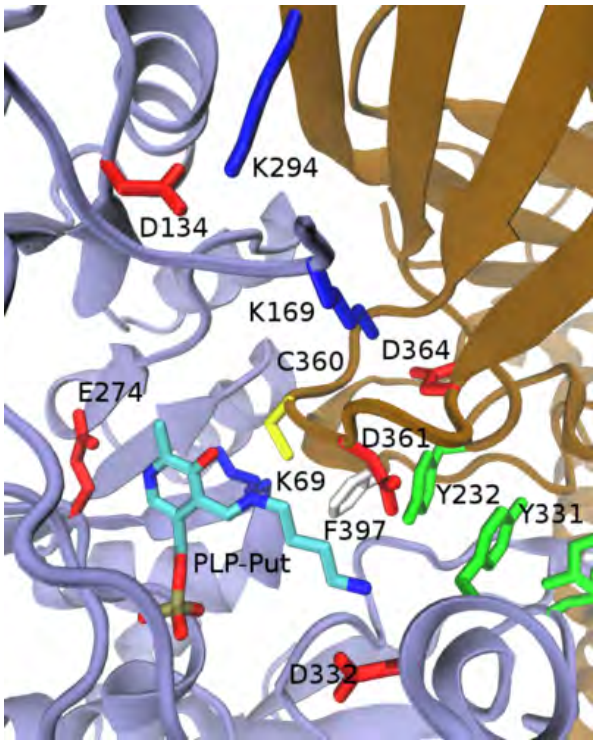


Figure 3.4: Active site and dimer interface residues in ornithine decarboxylase from *T. brucei* with the PLP-Putrescine (PLP-Put) adduct. Both ODC monomers (blue and brown, respectively) contribute residues to the active site. Acidic (red), basic (blue), cystien (yellow) residues are indicated The specificity 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).

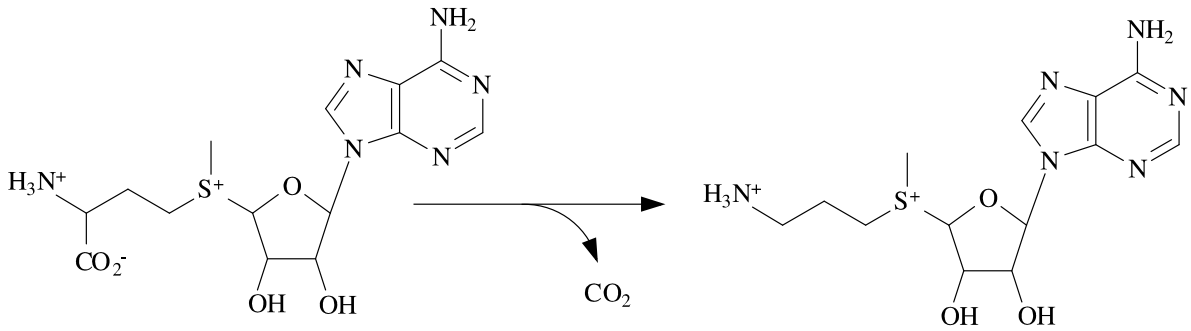


Figure 3.6: *S*-Adenosylmethionine decarboxylase reaction.

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).

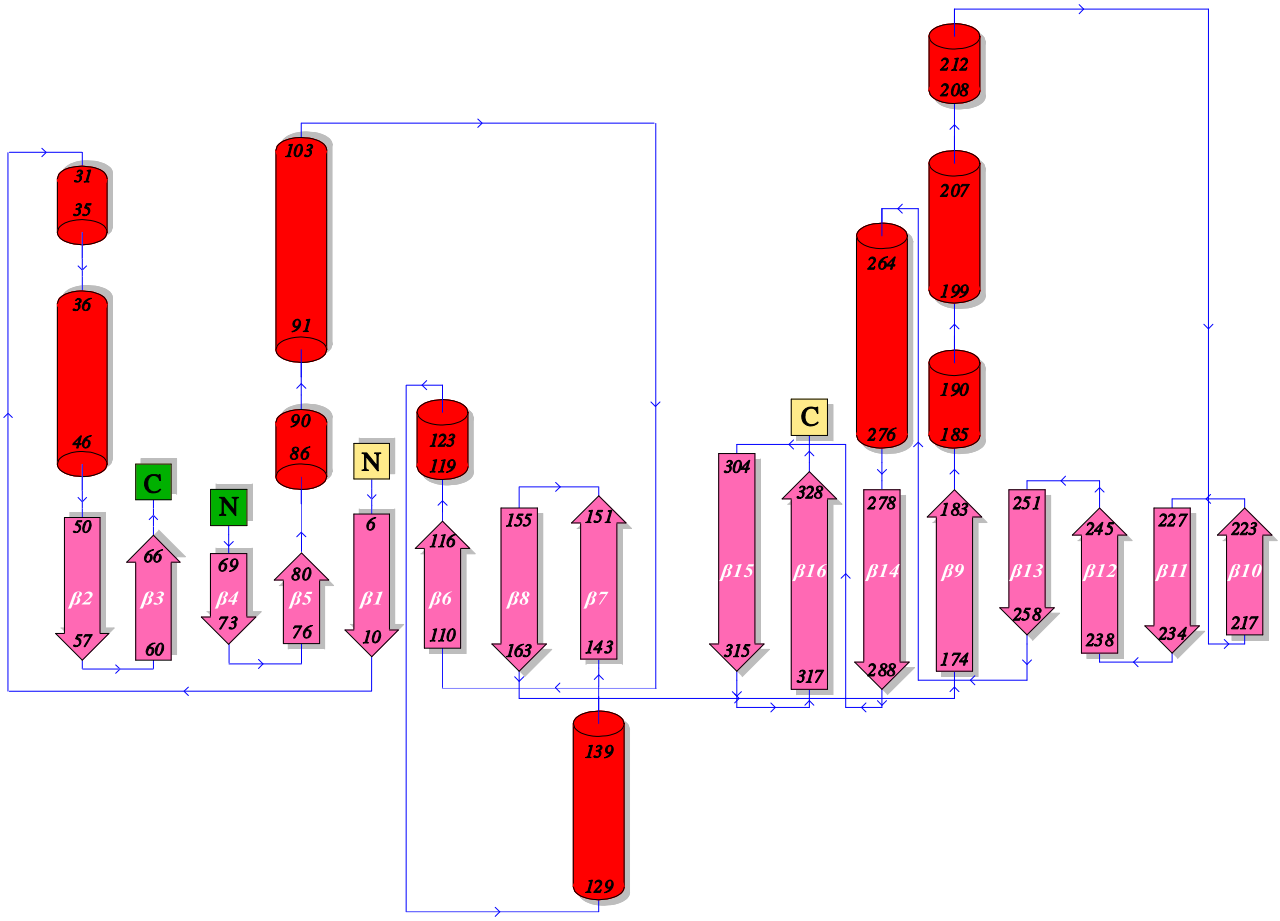


Figure 3.7: Topology of eukaryotic (human) AdoMetDC, PDB id: 1JEN. The site of internal proteolytic cleavage is indicated in green. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between 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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

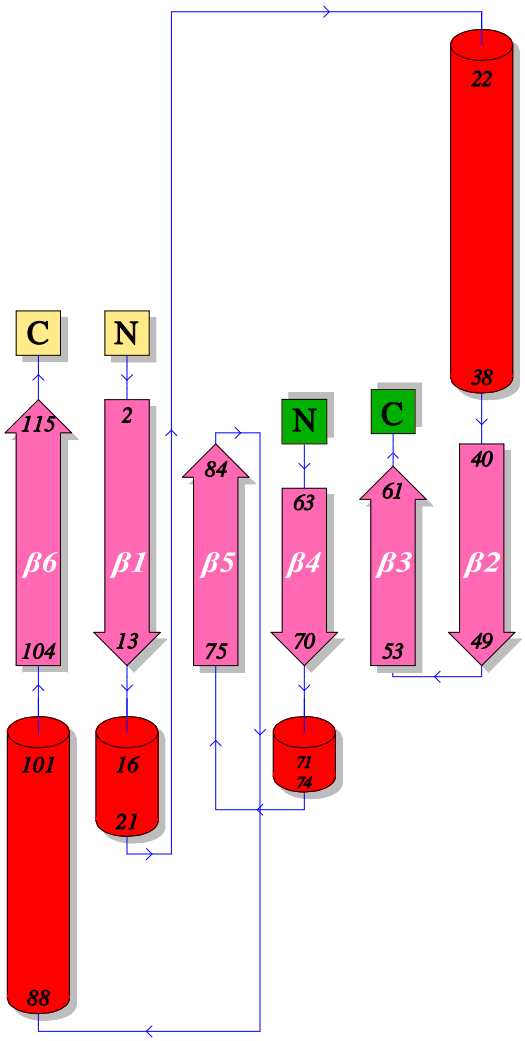


Figure 3.8: Topology of bacterial AdoMetDC from *Thermotoga maritima*, PDB id: 1TLU. Residue numbers for secondary structure element termini is included. β-sheets are numbered as indicated. Helix numbers are omitted due to the variable nature of these between species. The figure was adapted from that 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).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

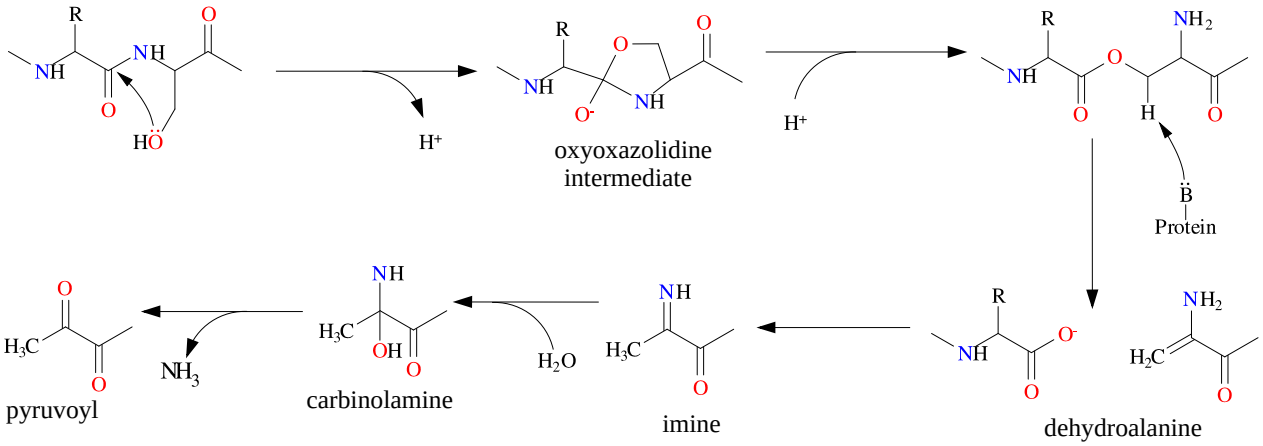


Figure 3.9: Pyruvoyl generation mechanism of *S*-adenosylmethionine decarboxylase takes place via an internal non-hydrolytic serinolysis. See text for further details. Adapted from Tolbert *et al.* (2003).

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).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

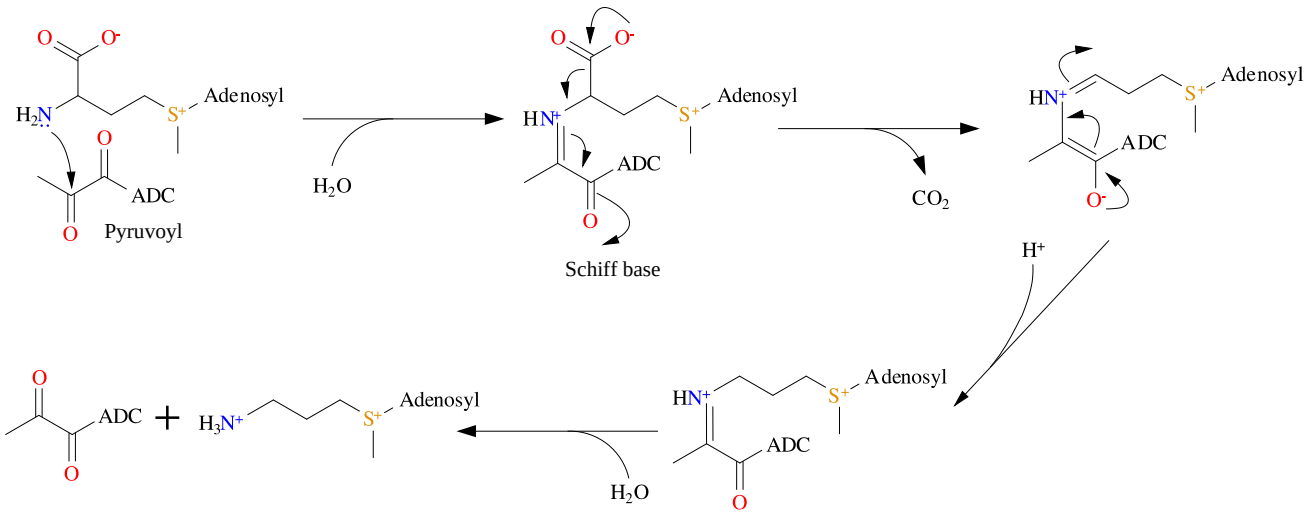


Figure 3.10: Reaction mechanism of *S*-adenosylmethionine decarboxylase. The pyruvoyl residues acts in much the same way as PLP-based decarboxylases. Formation of a Schiff base facilitates cleavage of the α-carboxylate group and CO_2. See text for further details. Adapted from Tolbert *et al.* (2001).

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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).

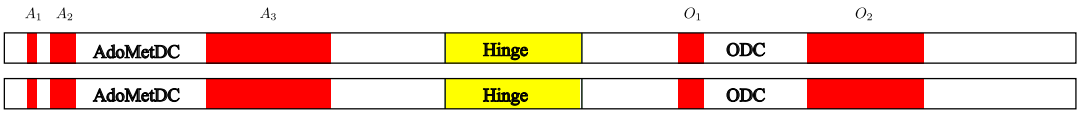


Figure 3.12: Linear structure of the *P. falciparum* AdoMetDC/ODC bifunctional complex. The domains, parasite-specific inserts (red) and hinge (yellow) regions are indicated.

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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

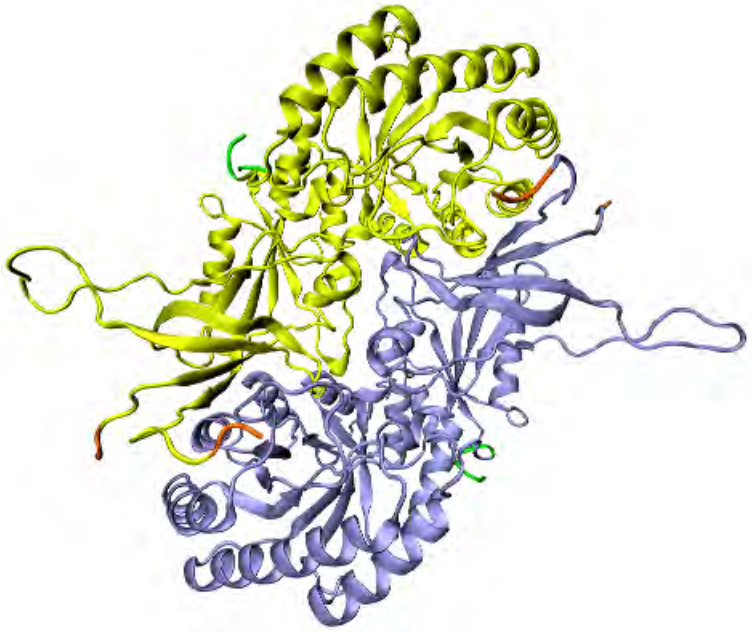


Figure 3.13: Homology model of *P. falciparum* ODC. Yellow, blue: monomers. Parasite-specific inserts are indicated in green (O_1) and orange (O_2). These inserts were too long to be modelled *ab initio* and most residues were thus excluded.

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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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.

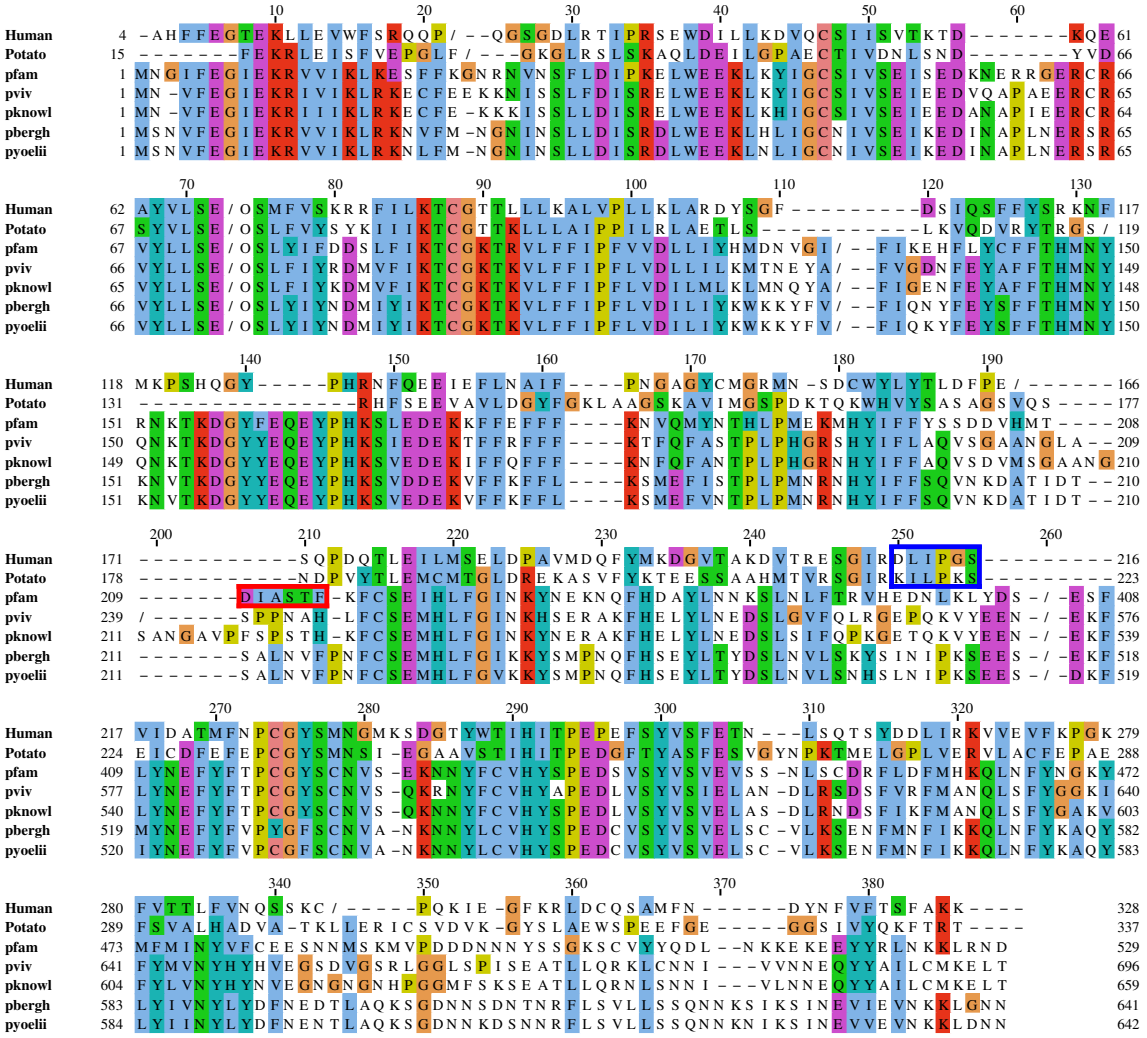


Figure 3.15: Alignment for modeling AdoMetDC. Human and *S. tuberosum* (potato) were used as templates, *P. berghei* (bergh) *P. falciparum* (pfam) *P. knowlesi* (pknowl) *P. vivax* (pviv) *P. yoelii* (pyoelii). The groups with red and blue borders were initially aligned one residue upstream (Wells *et 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.

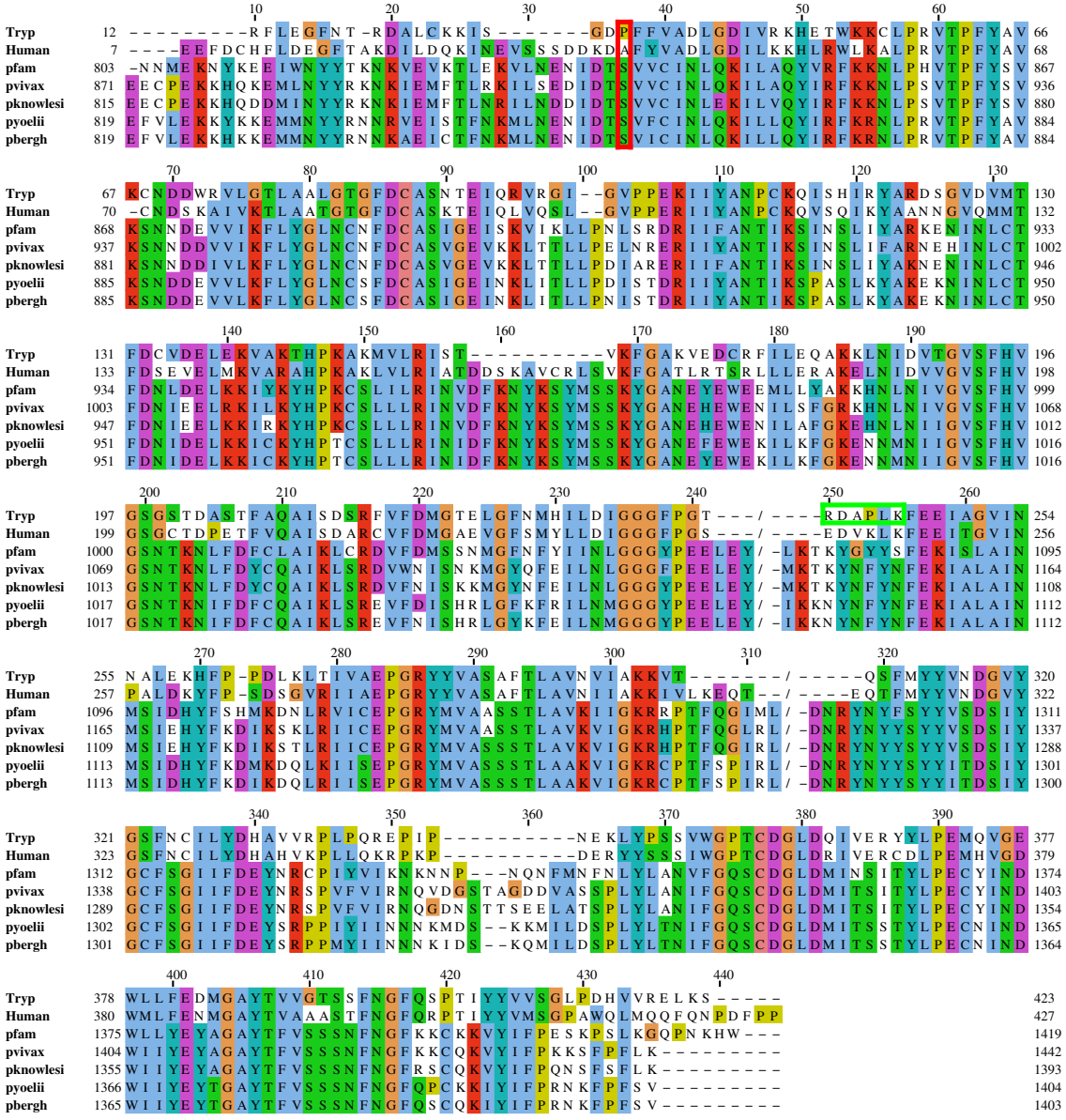
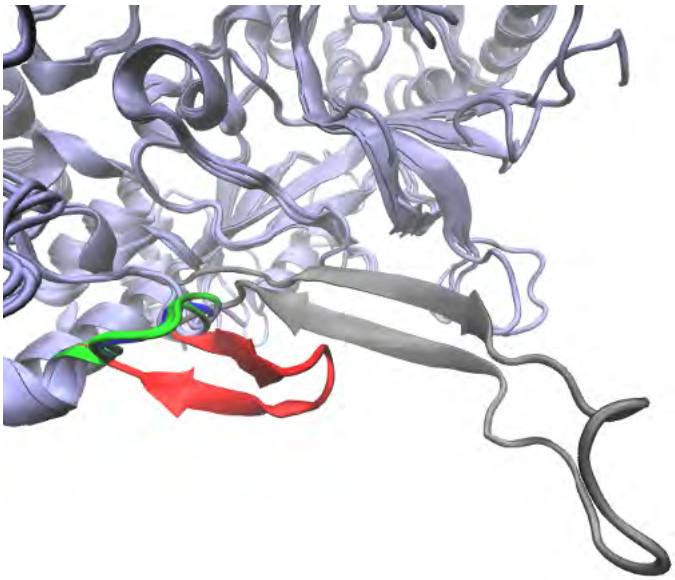
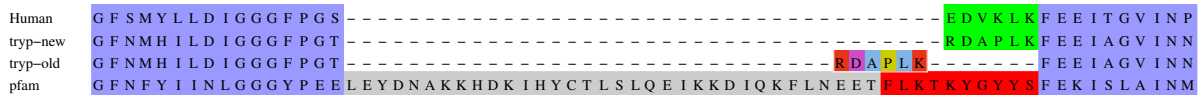


Figure 3.16: Alignment for modeling ODC. The column with the red border indicates the original position of the first residue used for modeling. The group in the green border was initially aligned eight residues upstream (Birkholtz *et al.*, 2003), and subsequently re-adjusted to make the first parasite-specific 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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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.

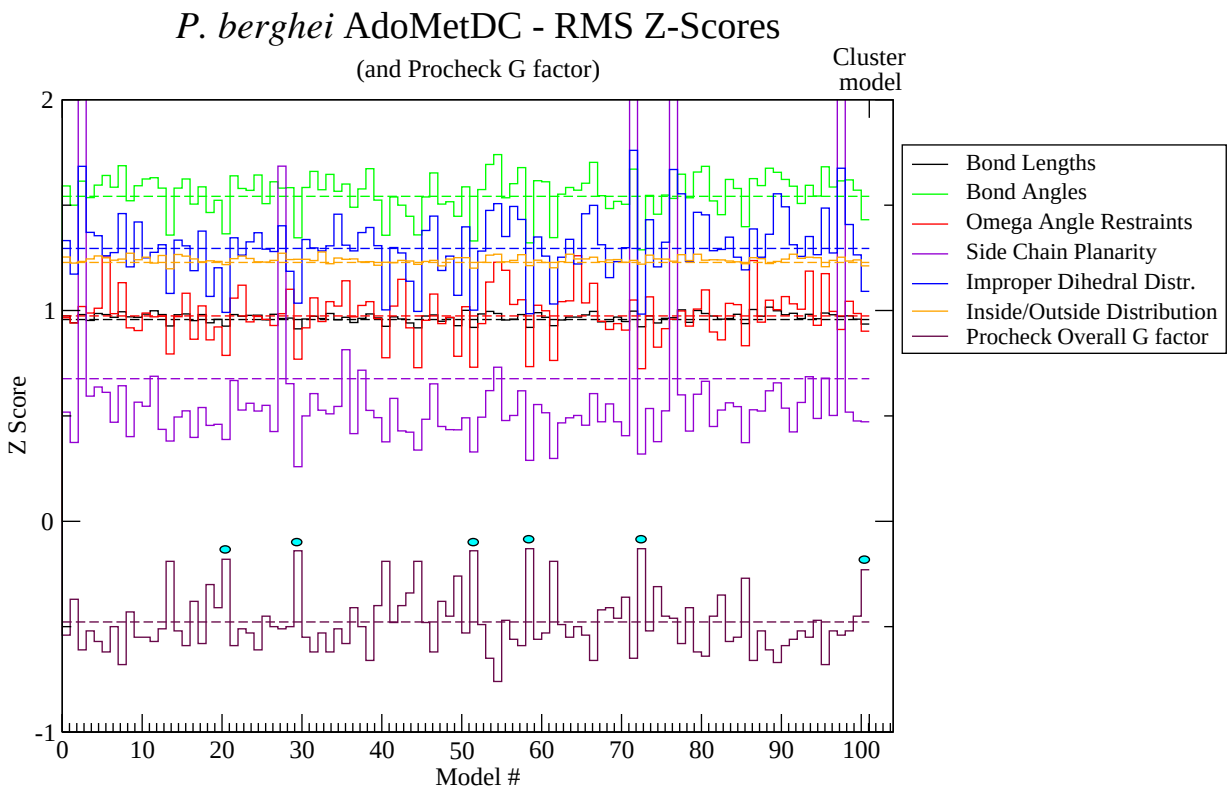


Figure 3.18: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. berghei* AdoMetDC. The models chosen for docking are indicated by the cyan dots. Averaged values are indicated by the dashed lines.

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

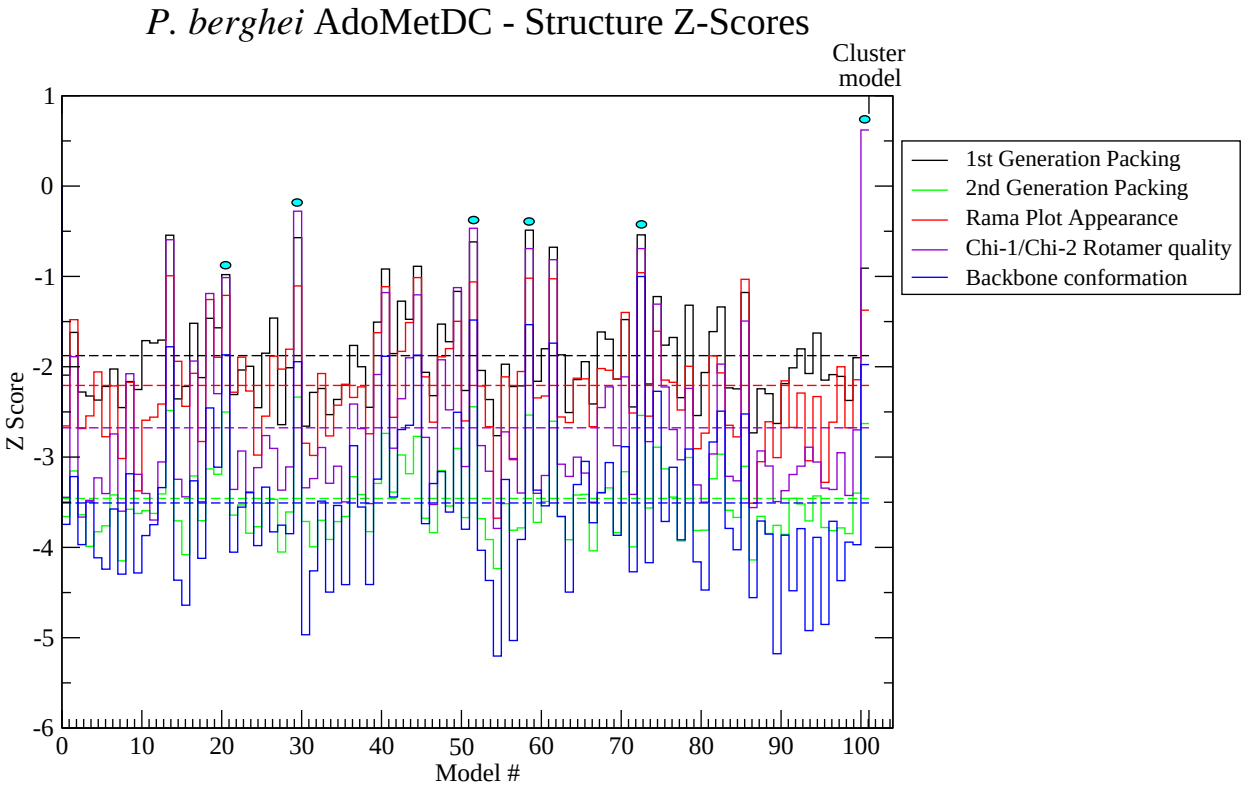


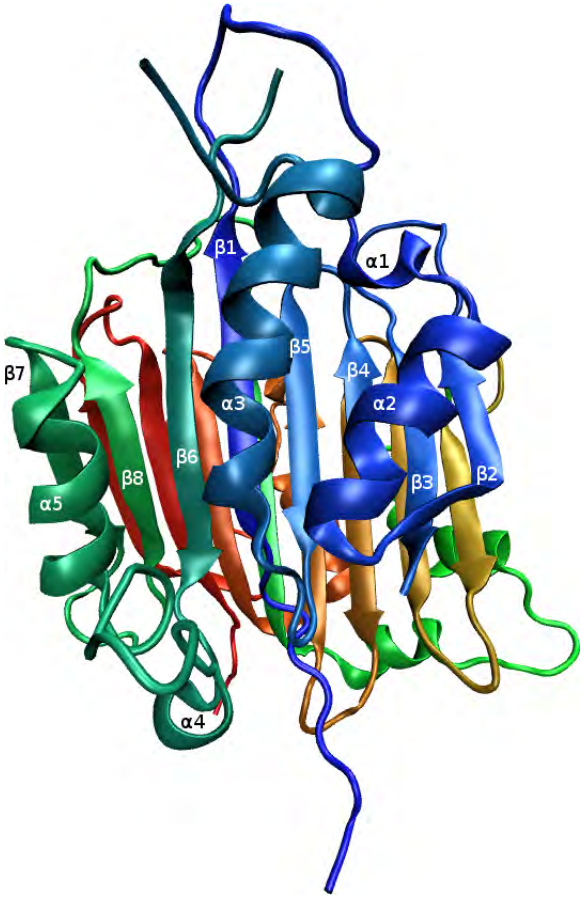
Figure 3.19: WHATIF Structure Z-scores for modeling of *P. berghei* AdoMetDC. The models chosen for docking are indicated by the cyan dots. Averaged values are indicated by the dashed lines.

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.

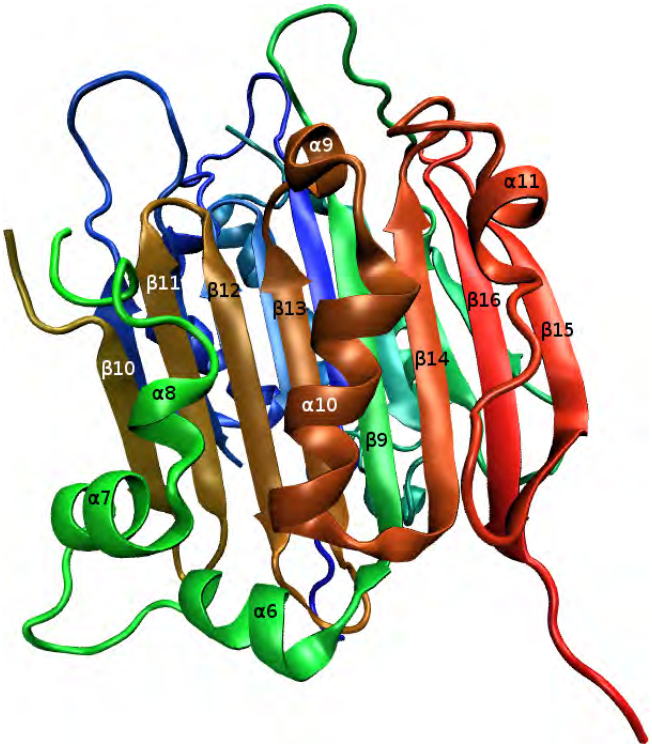
Species	AdoMetDC	ODC
P. berghei	21, 30, 52, 59, 73	6, 13, 23, 32, 58
P. falciparum	17, 32, 56, 57, 95	12, 18, 23, 44, 75
P. knowlesi	1, 3, 13, 32, 39	26, 32, 39, 60, 76
P. vivax	23, 39, 54, 69, 84	3, 43, 54, 56, 75
P. yoelii	4, 47, 59, 63, 69	6, 58, 76, 78, 98

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.



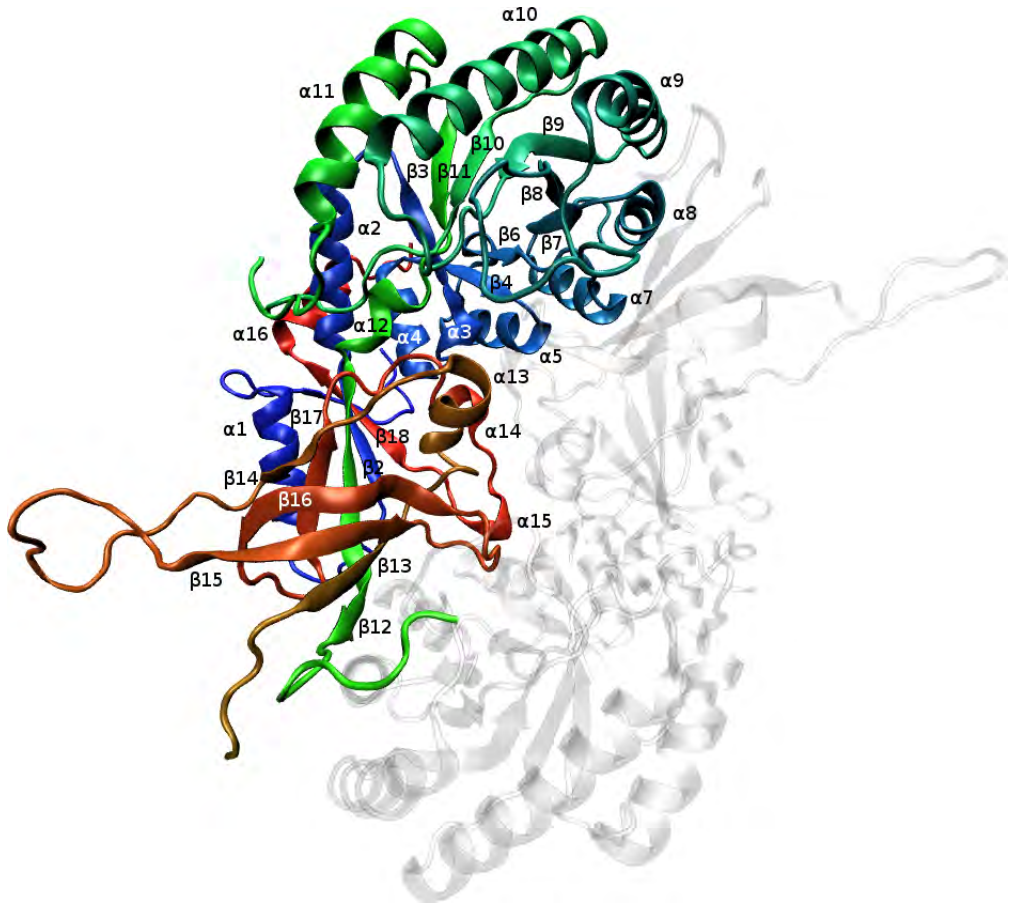
(a) Top



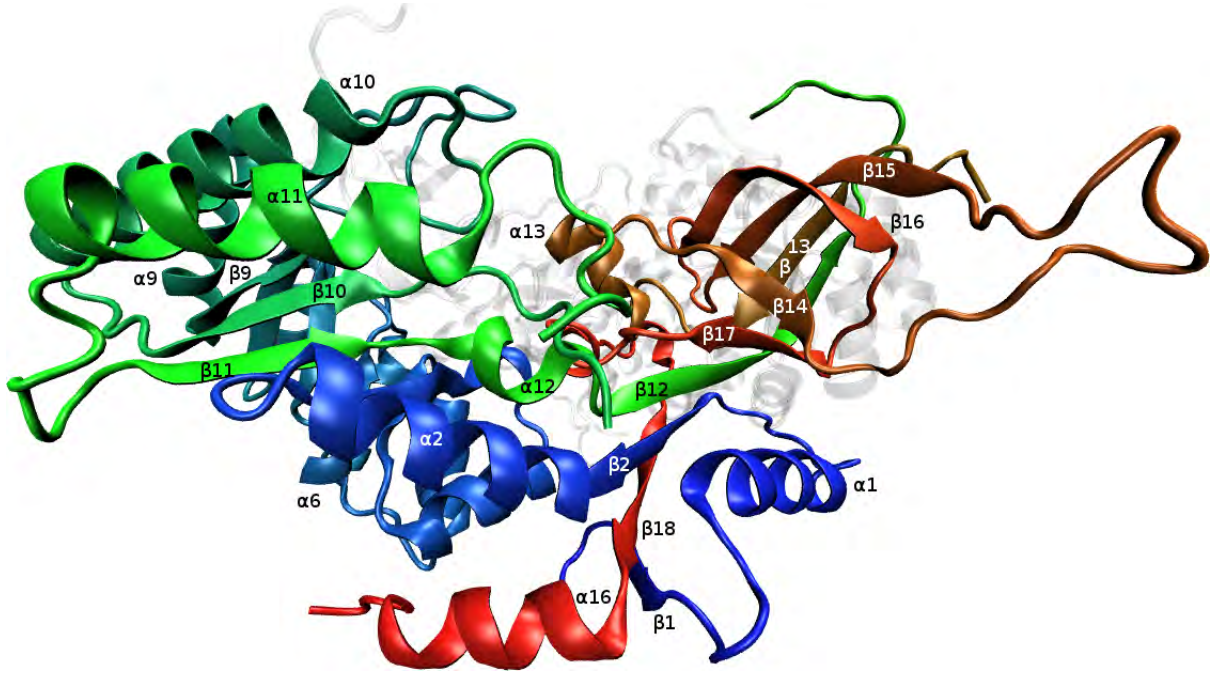
(b) Bottom

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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(a) Top

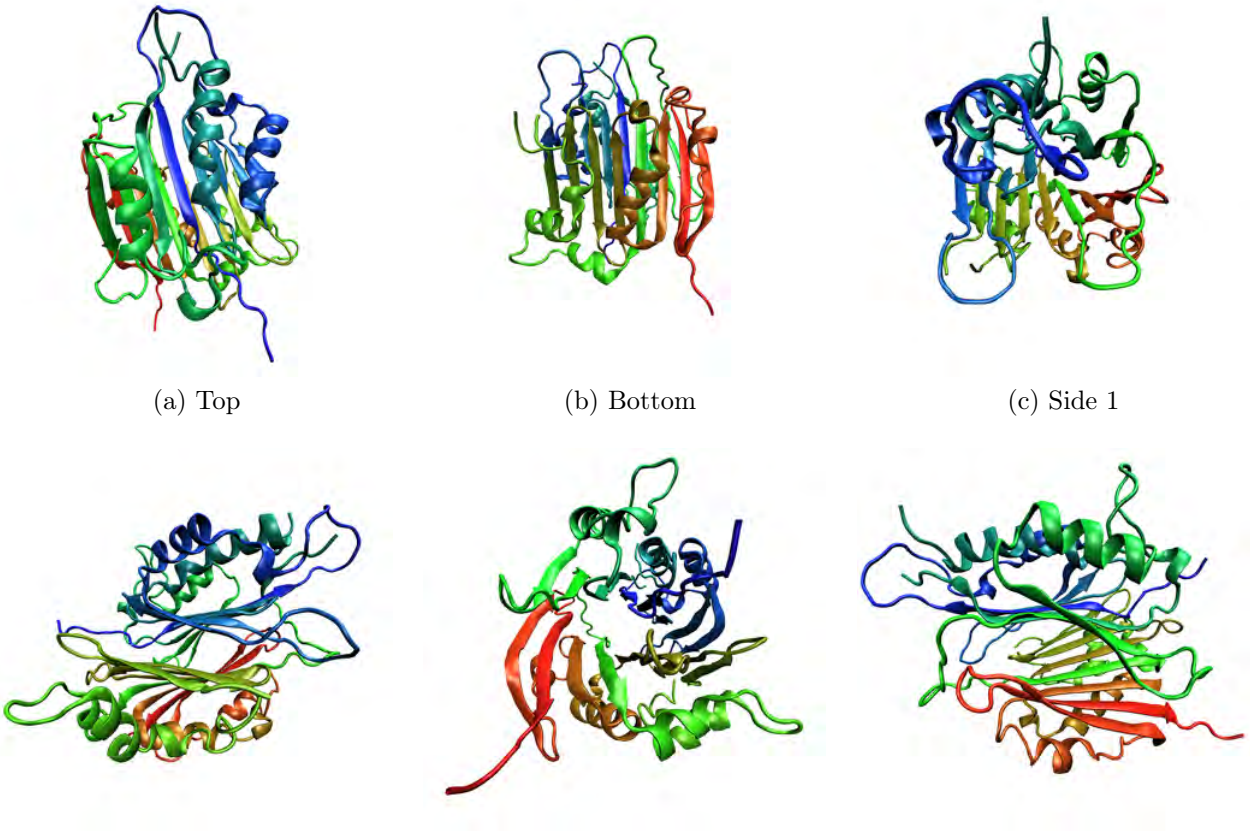


(b) Side

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

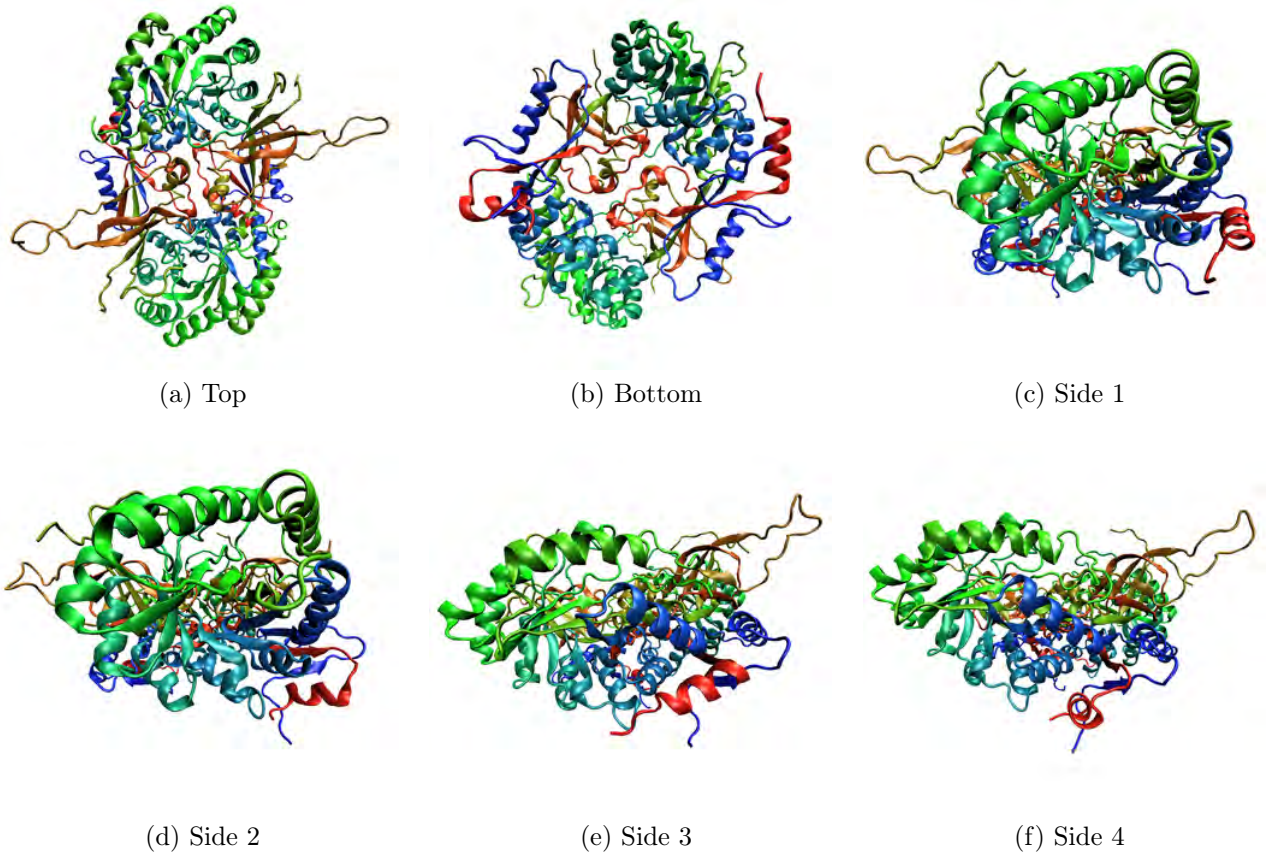
For the purpose of analysing docking results it is useful to define poses of both AdoMetDC and ODC that cover the entire surface of the respective protein or protein complex. For each domain six roughly orthogonal poses are arbitrarily defined (Fig. 3.22 and 3.23). While AdoMetDC is asymmetric, within 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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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.

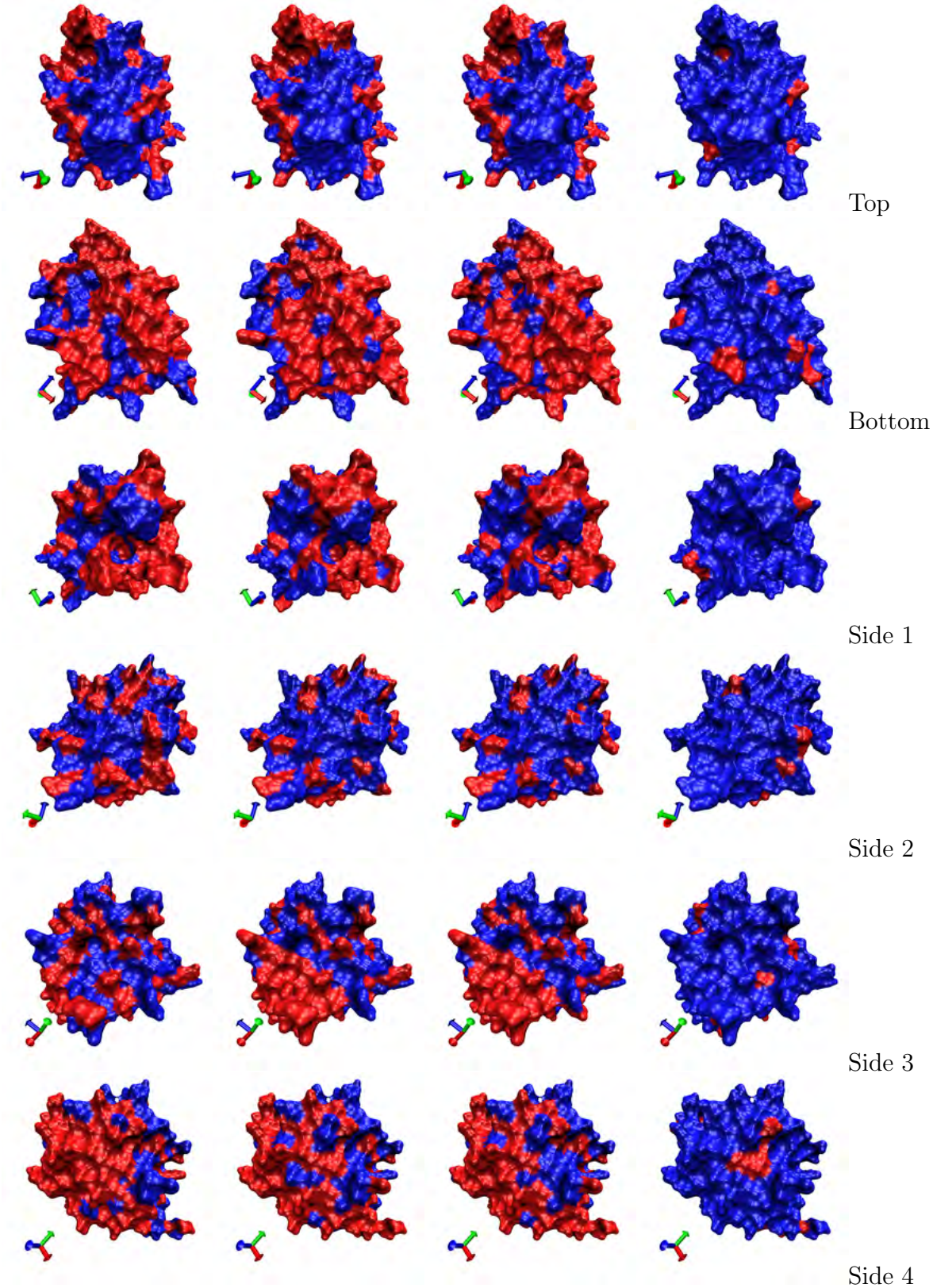


Figure 3.24: Pairwise conservation of *P. berghei* AdoMetDC surface residues. *P. berghei* is compared to *P. falciparum*, *P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1-6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: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

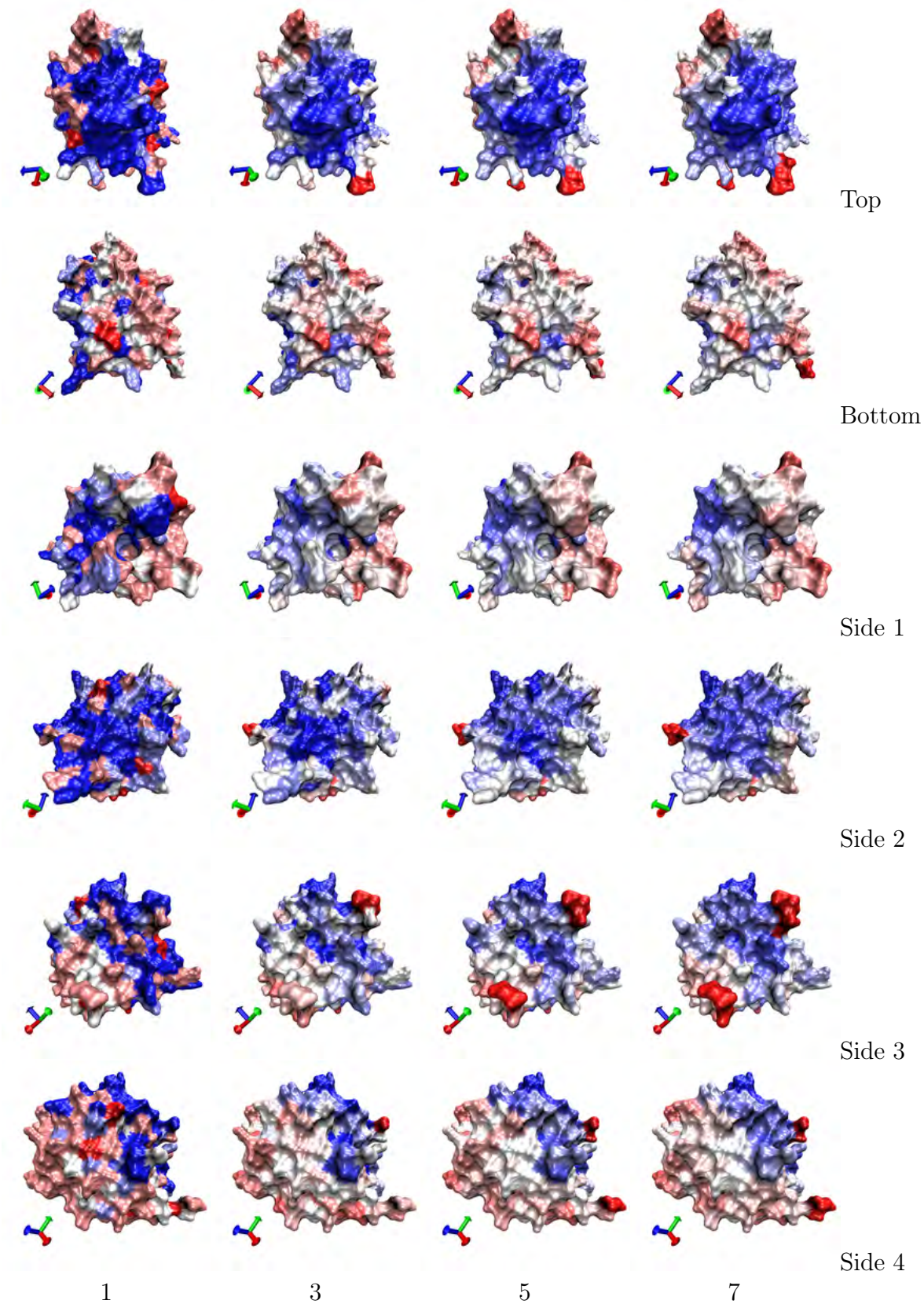


Figure 3.25: Conservation of *P. berghei* AdoMetDC surface residues. *P. berghei* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

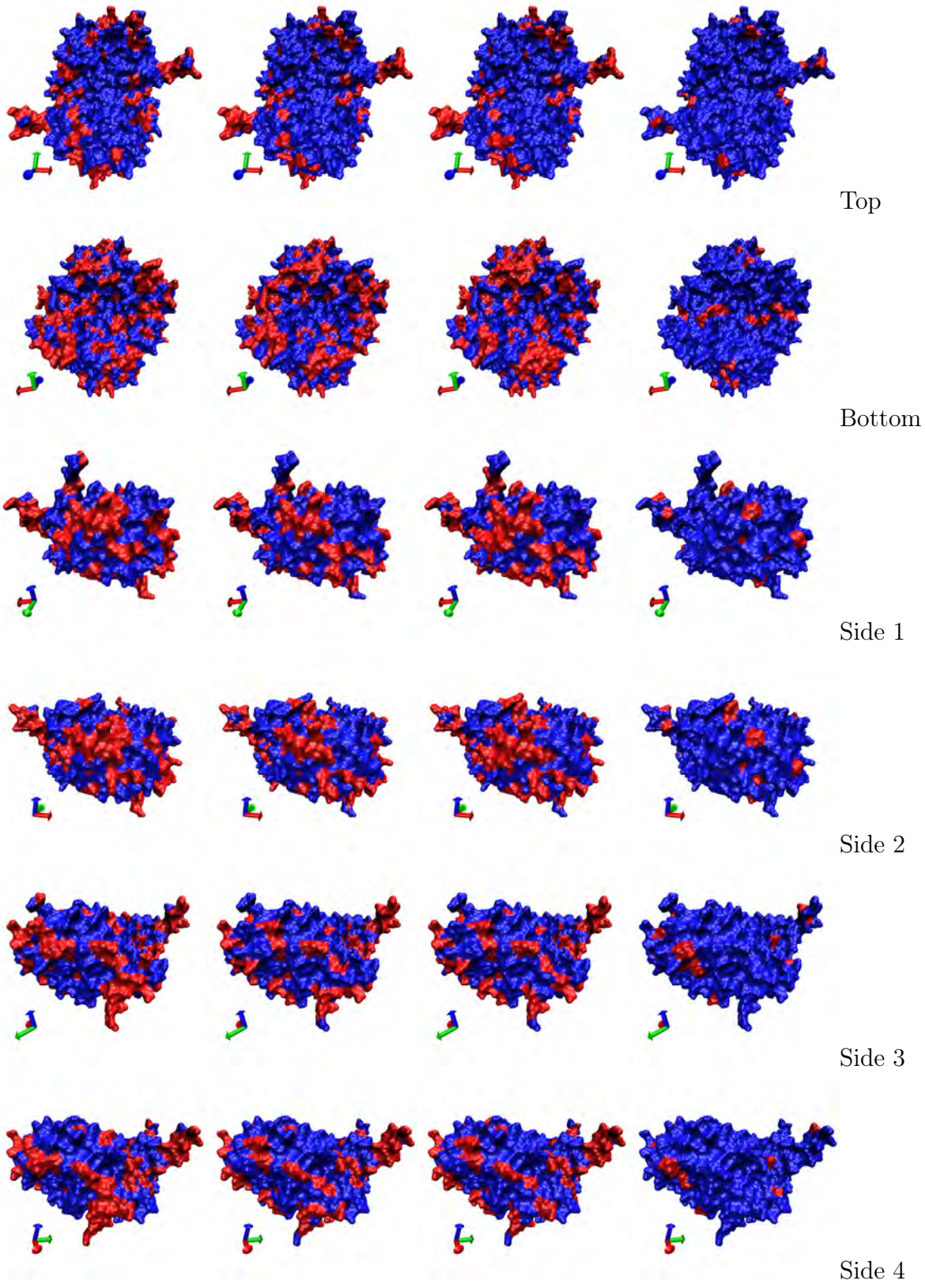


Figure 3.26: Pairwise conservation of *P. berghei* ODC surface residues. *P. berghei* is compared to *P. falciparum*, *P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: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

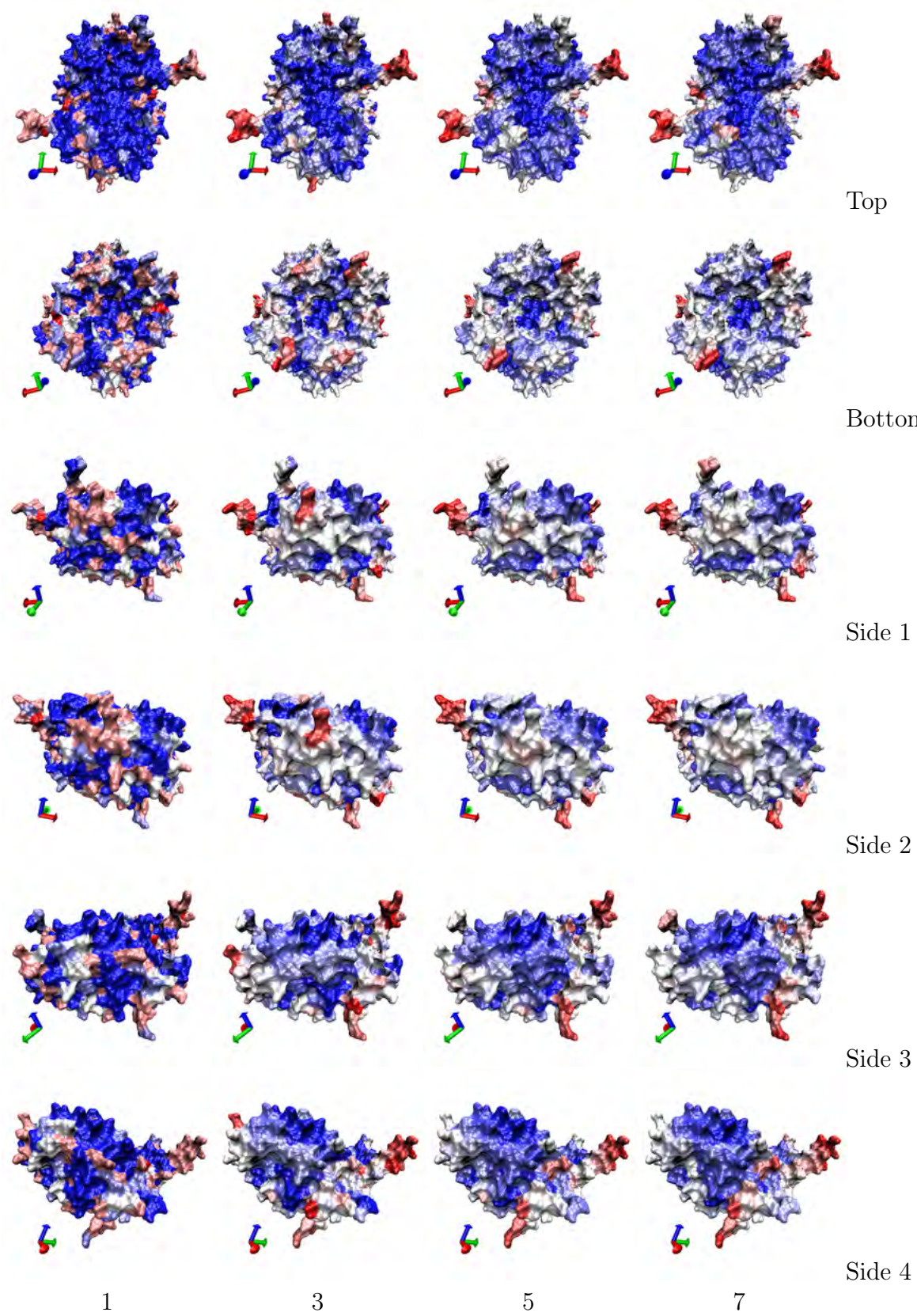


Figure 3.27: Conservation of *P. berghei* ODC surface residues. *P. berghei* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses,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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

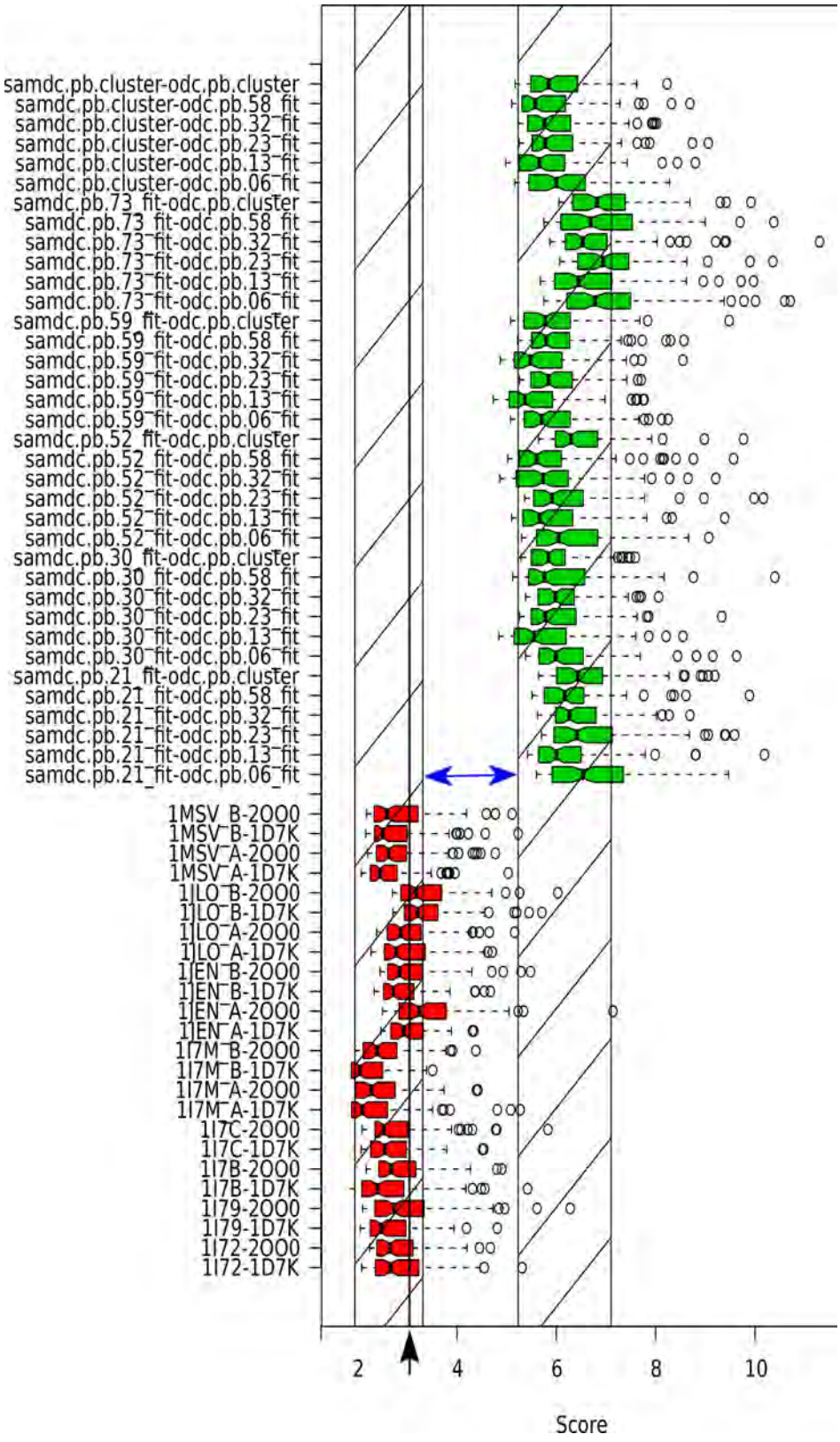


Figure 3.28: Distribution of top 100 RP scores for docking of human (red) and *P. berghei* (green) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human with a gap separating them (black arrow). The gap is too small to be visible on this scale. One 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).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.4.2.2 Centre of mass distribution

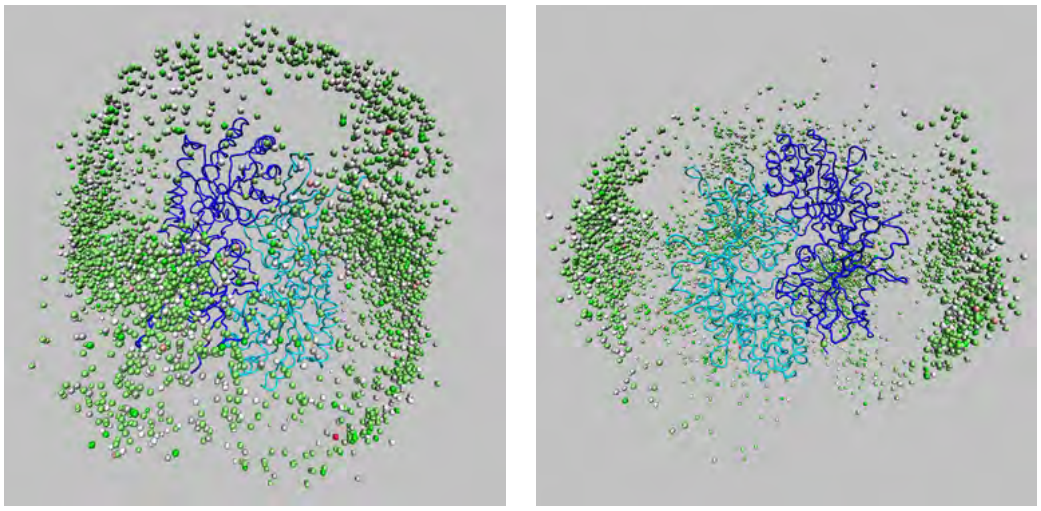
The distribution of the centre of mass (COM) of each domain was analysed relative to the partner protein. An uneven distribution was obtained for both AdoMetDC and ODC for the top 100 predictions across all five *Plasmodium* species. Certain trends were also observed when comparing the various species. 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.

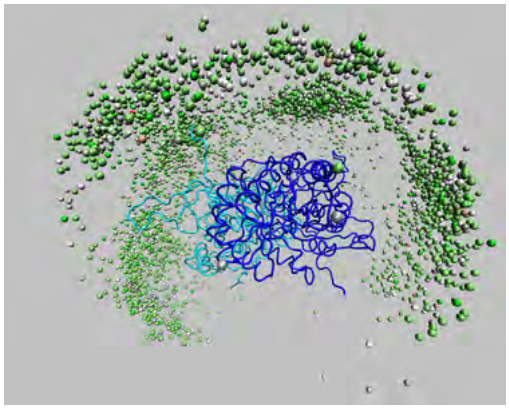
[3] *P. berghei* is included as representative of four out of the five species, while *P. falciparum* is slightly atypical and therefore discussed further here. The remaining results are included in the Appendix (B.4)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

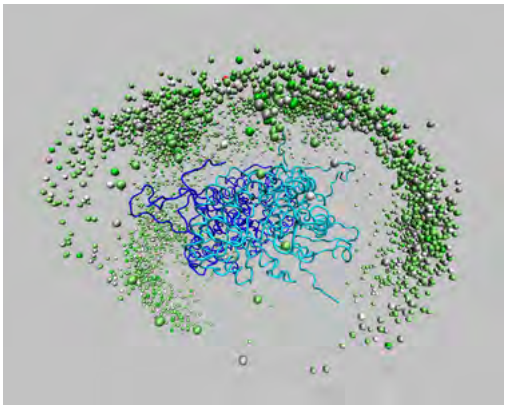


(a) Top

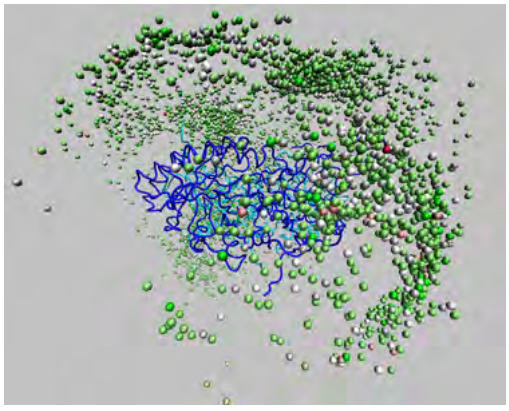
(b) Bottom



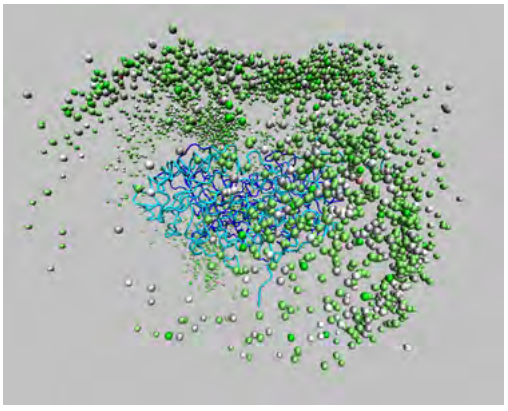
(c) Side 1



(d) Side 2

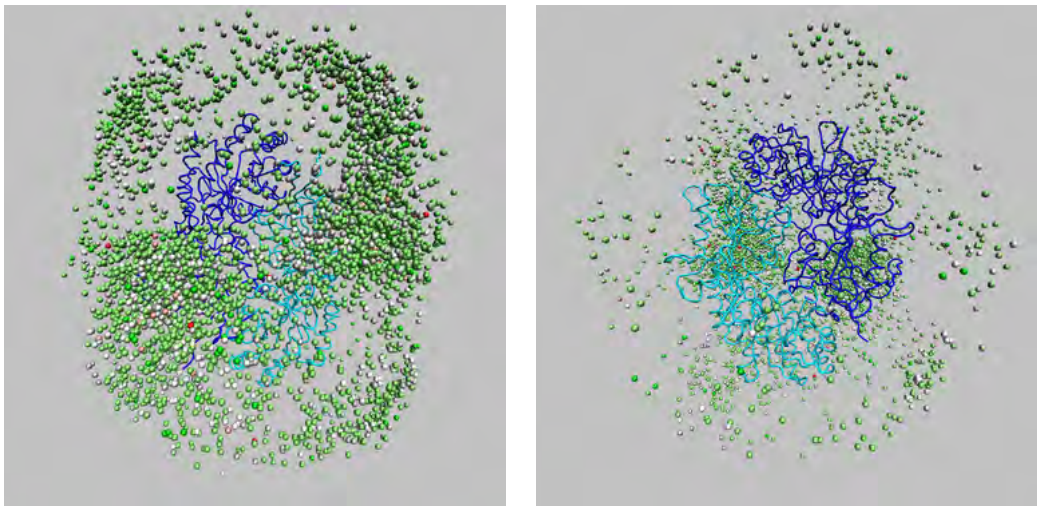


(e) Side 3



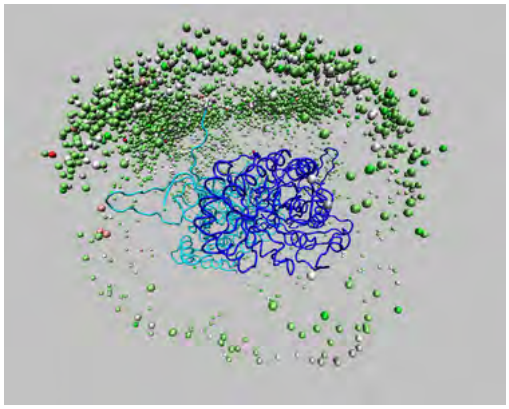
(f) Side 4

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).

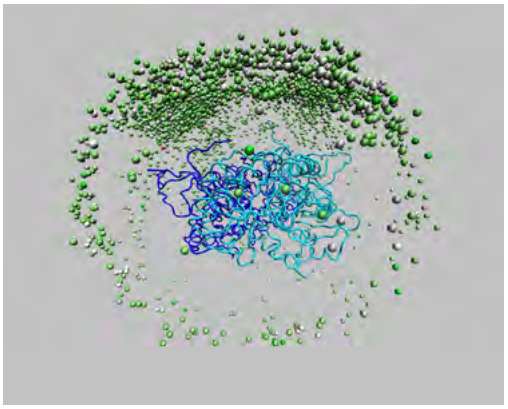


(a) Top

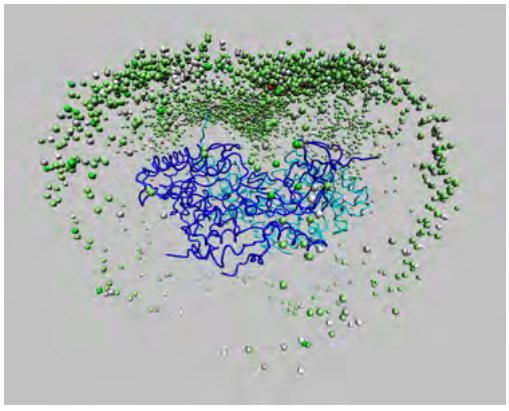
(b) Bottom



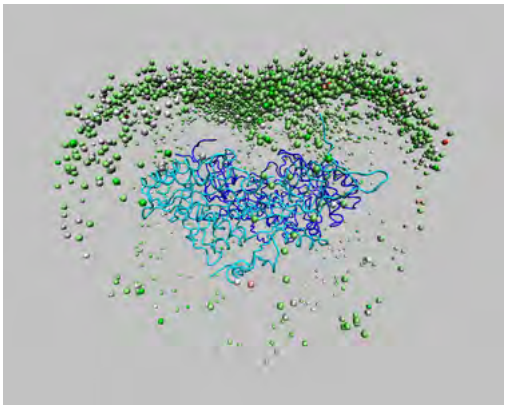
(c) Side 1



(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).

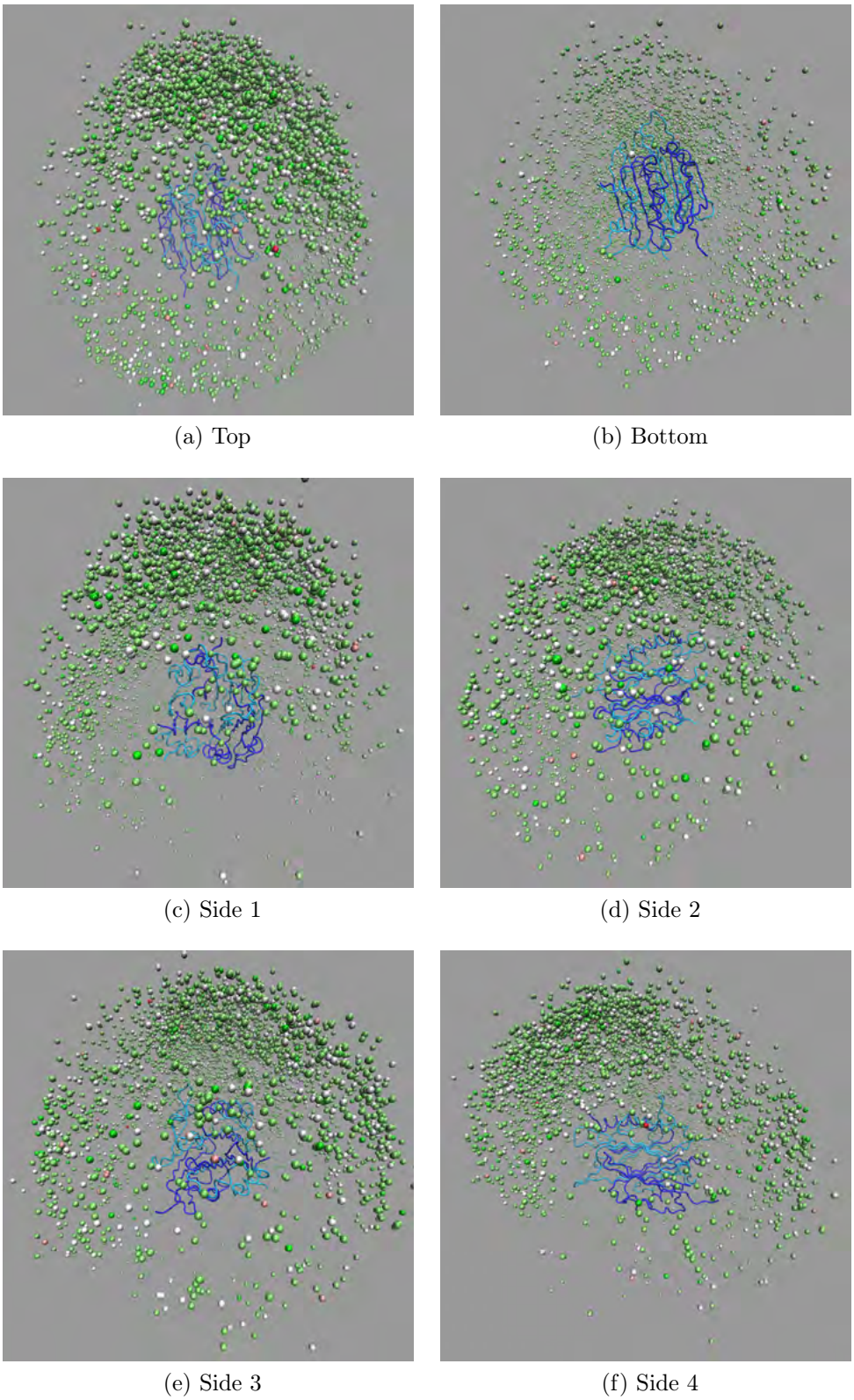
The distribution of ODC COMs relative to AdoMetDC was less consistent across the five *Plasmodium* species. The most consistent pattern, is again a favouring of the top face

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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)

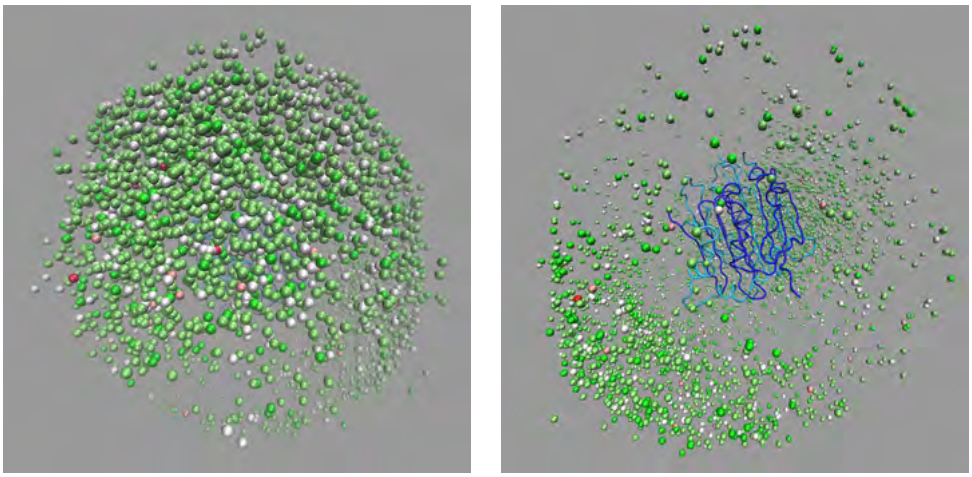


(a) Top (b) Bottom

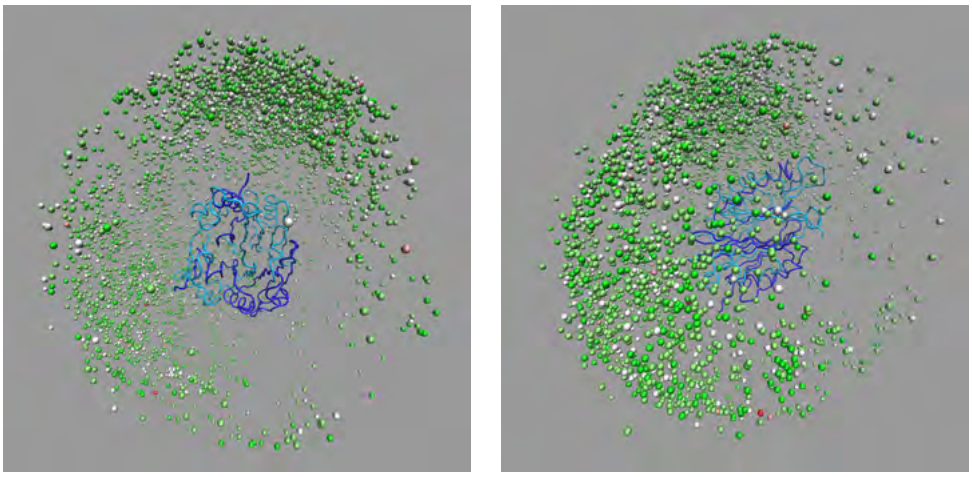
(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

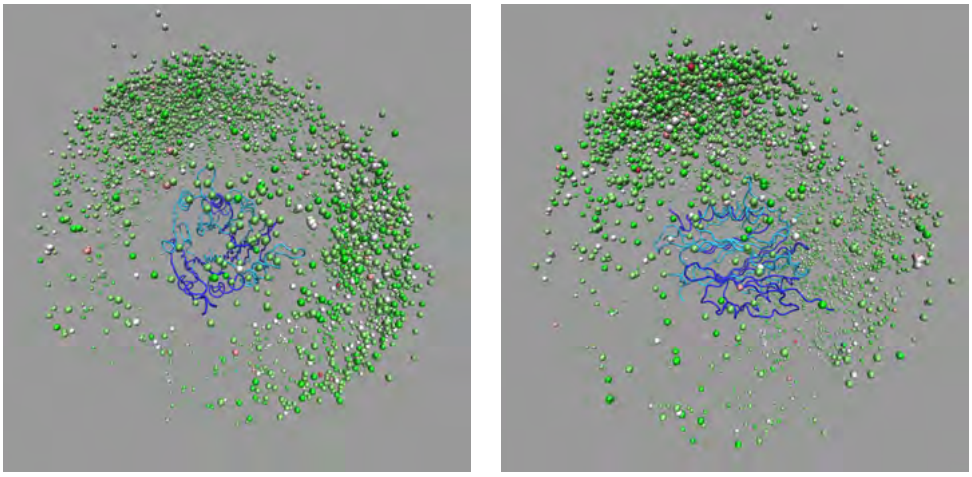
Figure 3.31: Centre of mass (COM) distributions of *P. berghei* ODC relative to AdoMetDC. AdoMetDC chains A & C (blue) and (cyan) B & D are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based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



(a) Top (b) Bottom



(c) Side 1 (d) Side 2

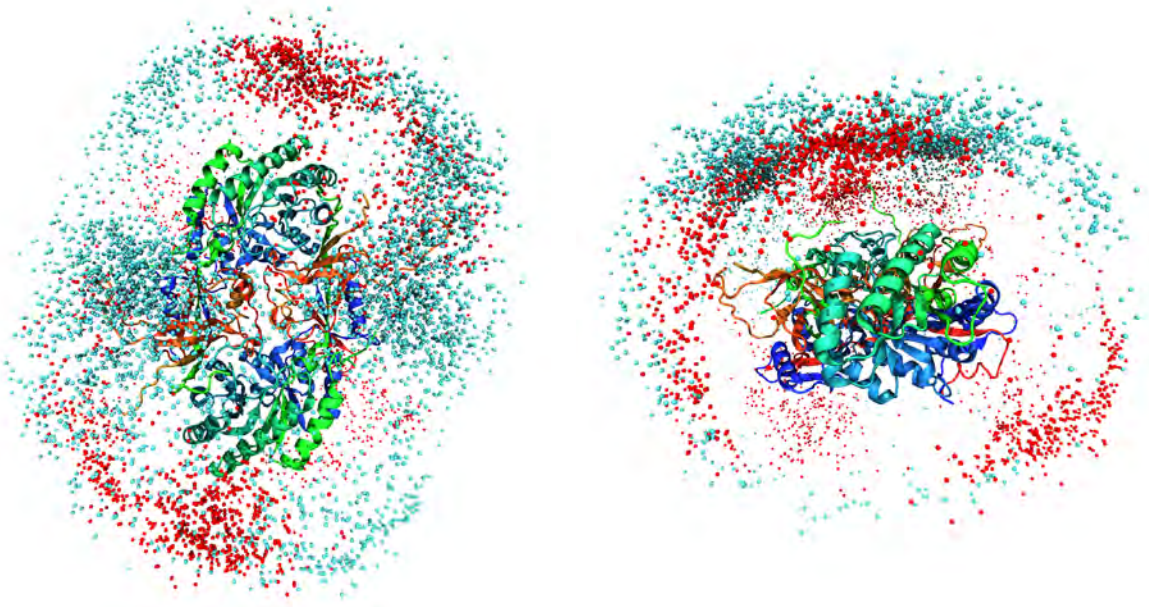


(e) Side 3 (f) Side 4

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).

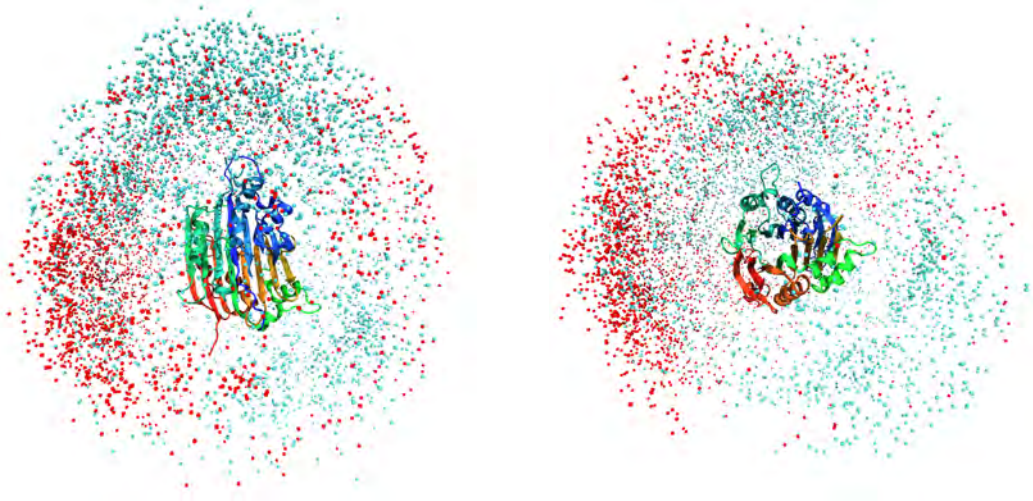


(a) Top view of human and *P. falciparum* AdoMetDC COMs

(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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



(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.

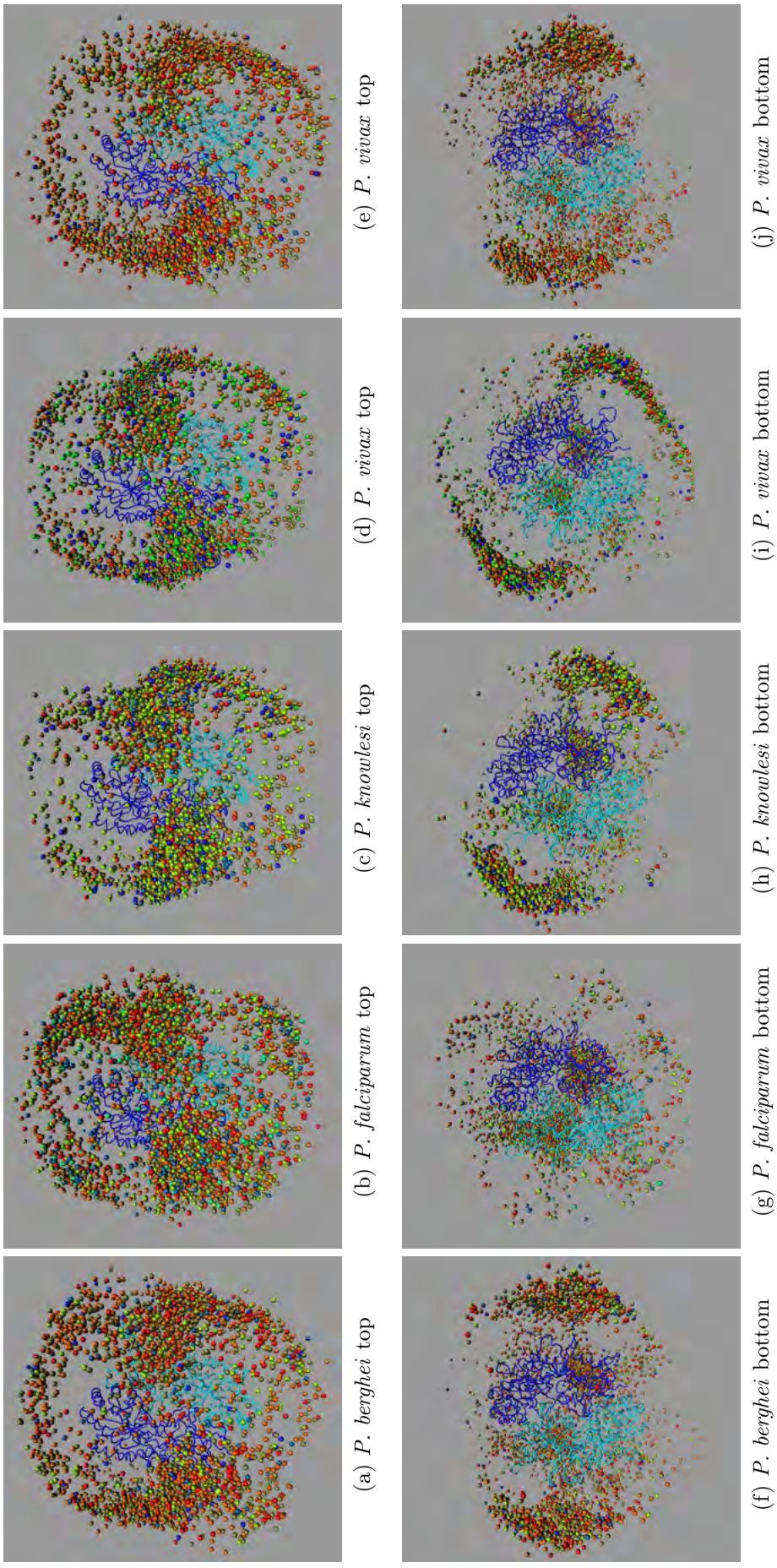
3.4.2.3 Mutually favoured contacting regions

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

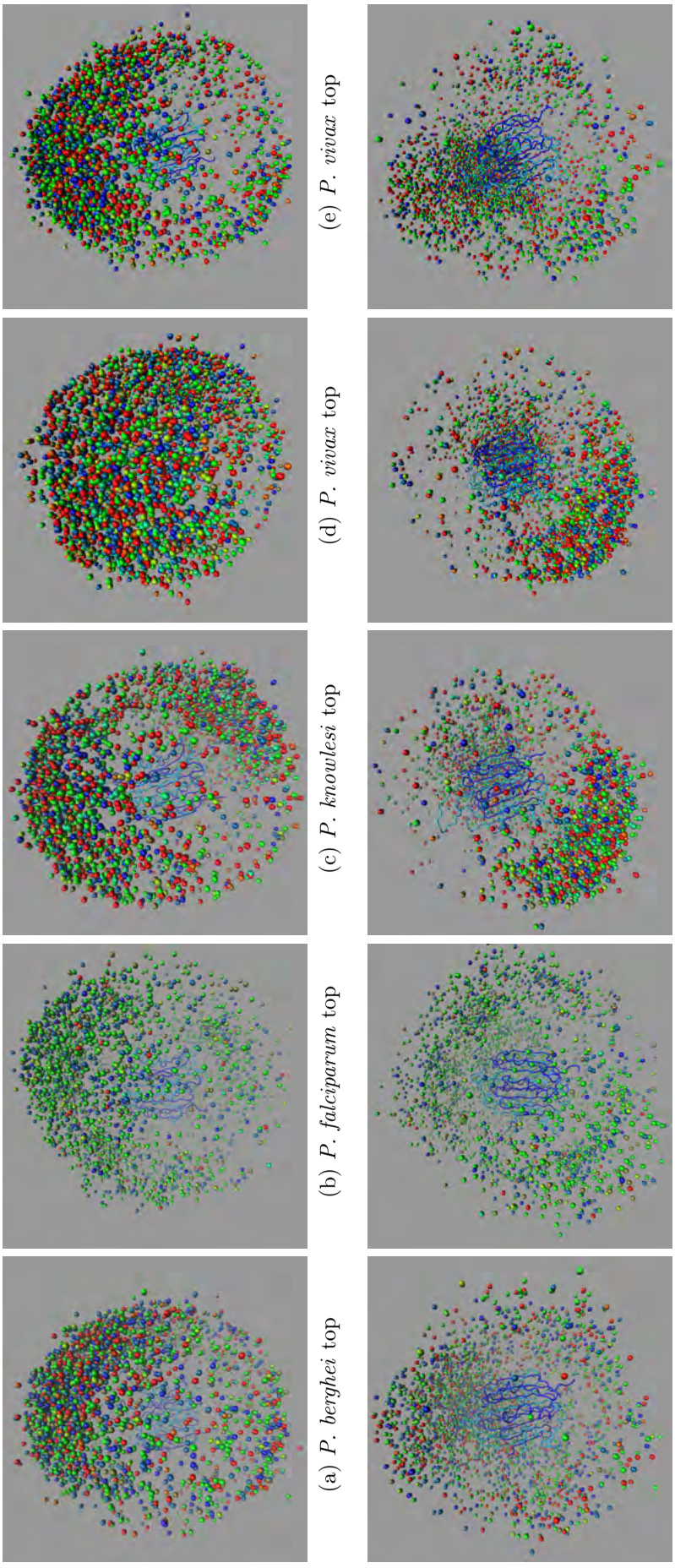
The data visualisation program OPENDX was used to construct heat-maps of the number of times a certain AdoMetDC/ODC residue pair made contact within the 3600 docking runs analysed. Heat-maps were constructed for each *Plasmodium* species. The distance chosen was 15 Å for the C_α^{ADC}–C_α^{ODC} distance.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

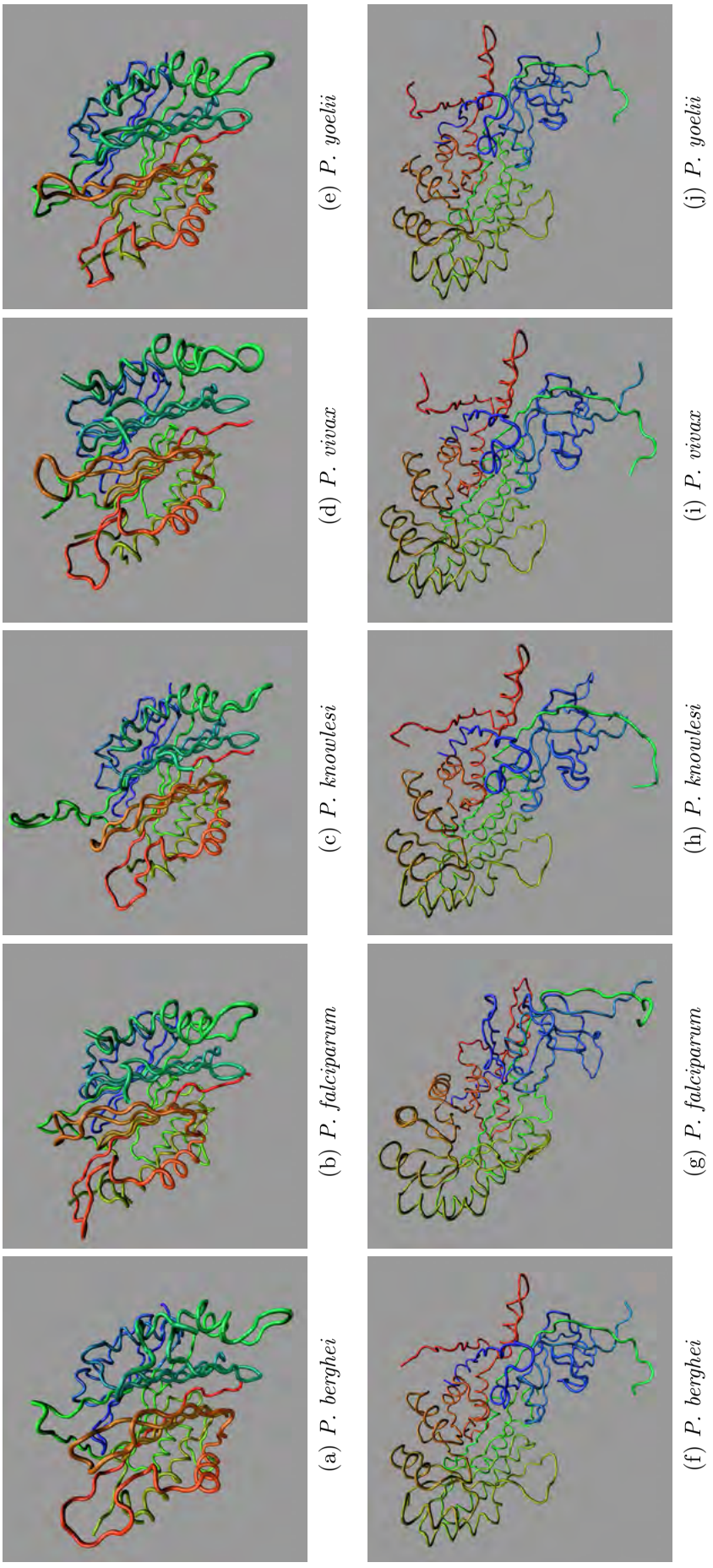
Figure 3.35: AdoMetDC COMs coloured according to closest contacting residue. Red → Green → Blue = increasing residue number (Fig. 3.37). There is no obvious clustering according to colour, i.e. there isn't obvious preference for AdoMetDC/ODC surface patches to associate.



(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.

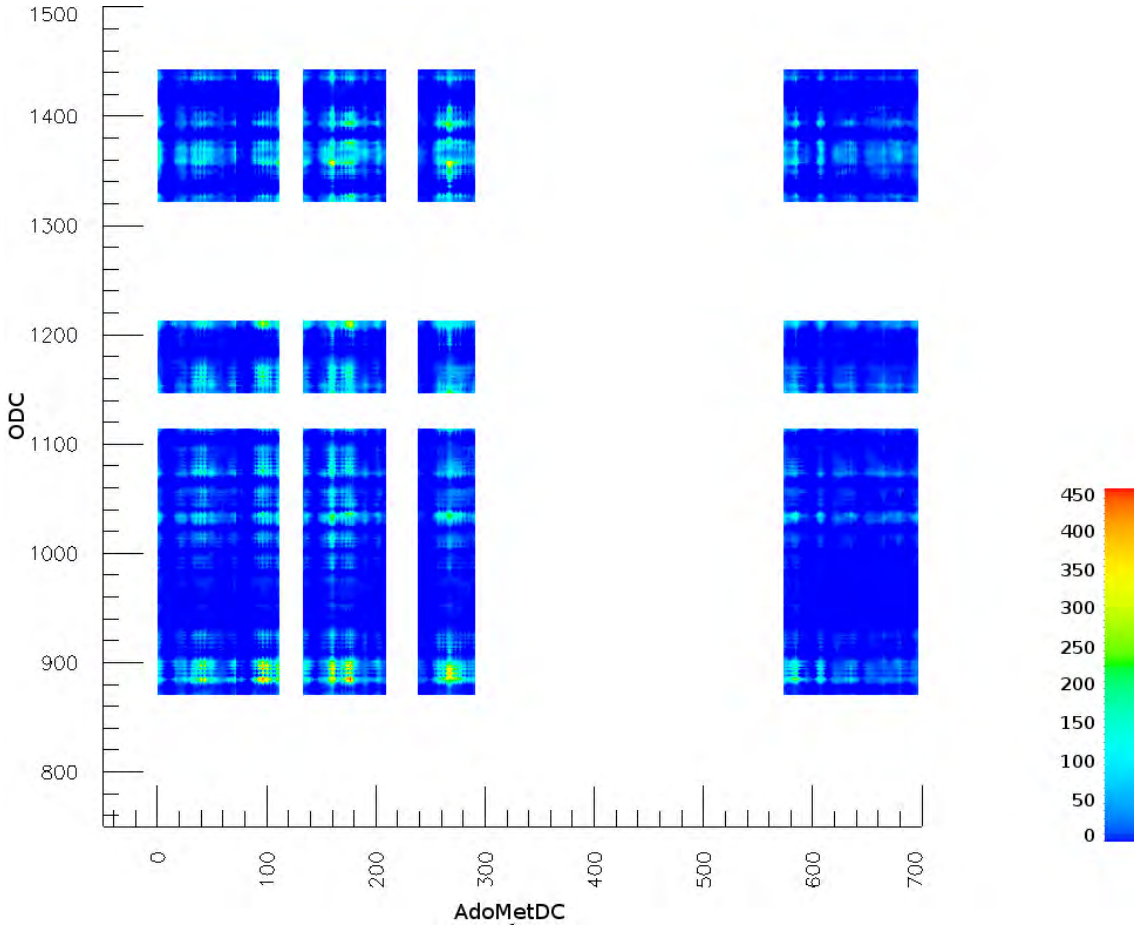


Figure 3.38: Contact count heat-map for *P. berghei*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_α^{ADC}-C_α^{ODC}<15 Å). The typical range is between 0-500,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

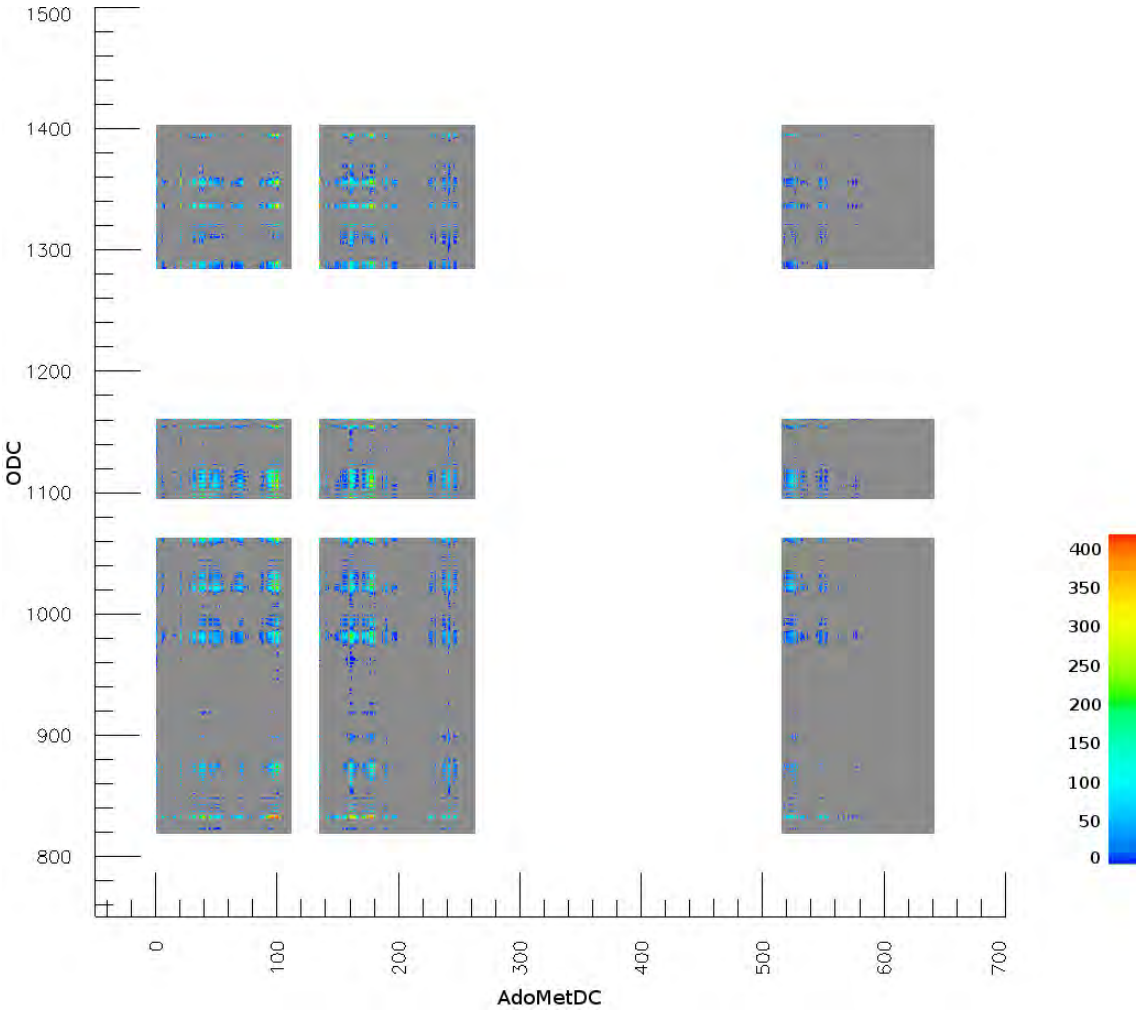
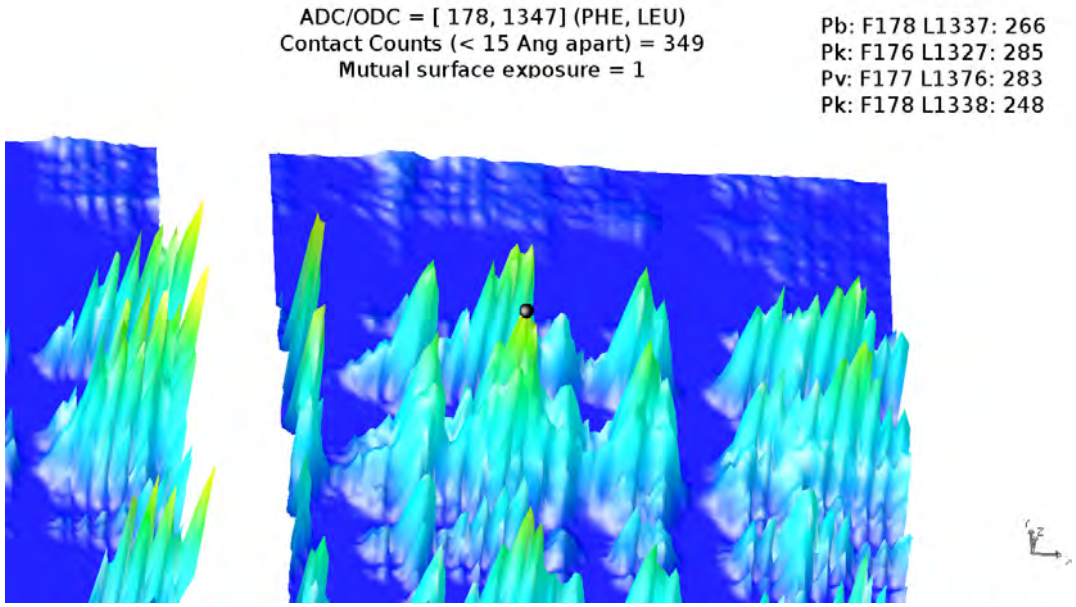
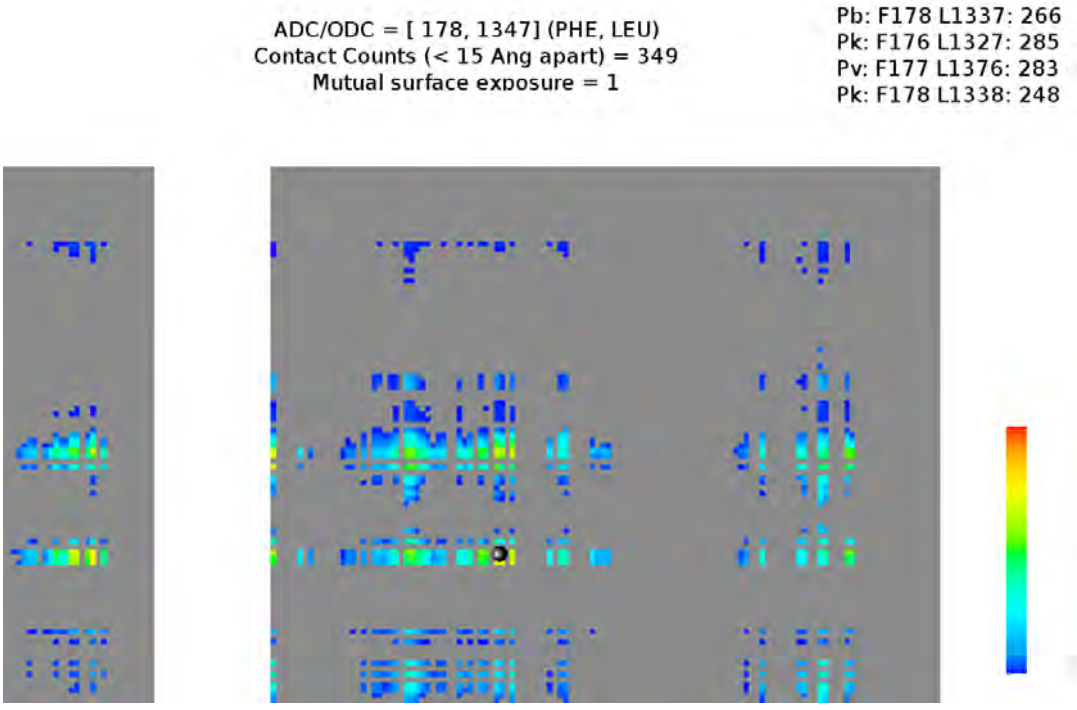


Figure 3.39: *P. berghei* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes contact (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.

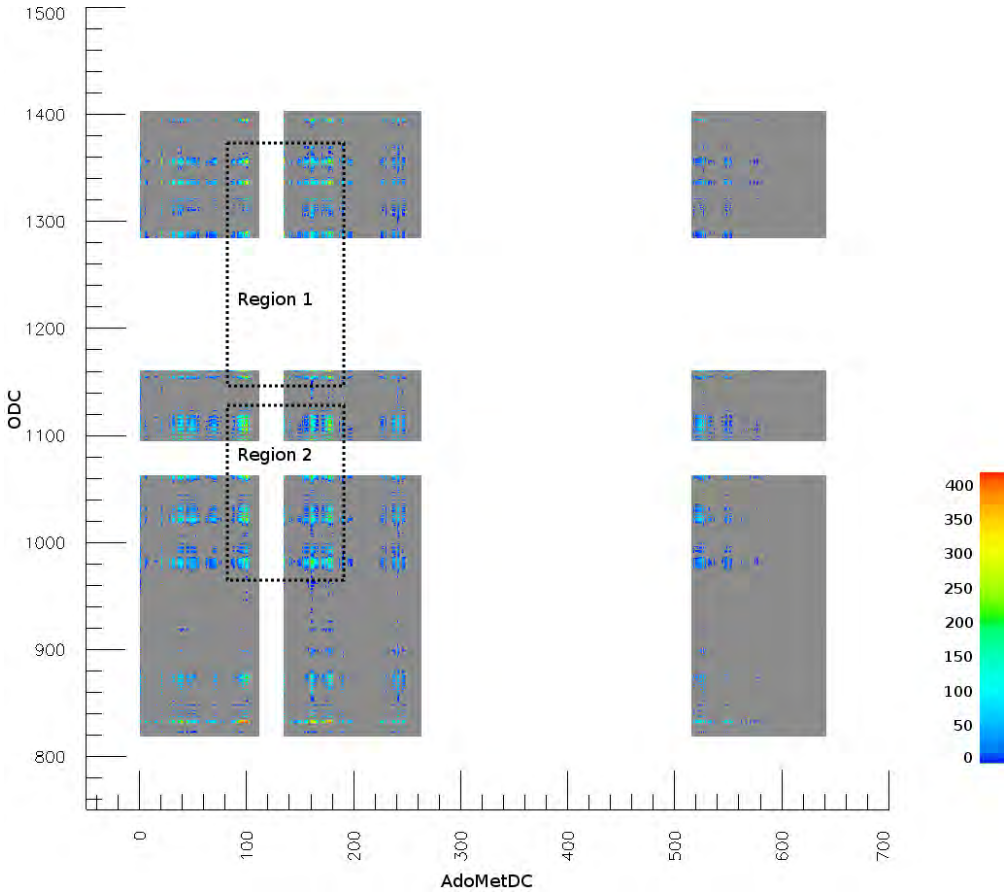
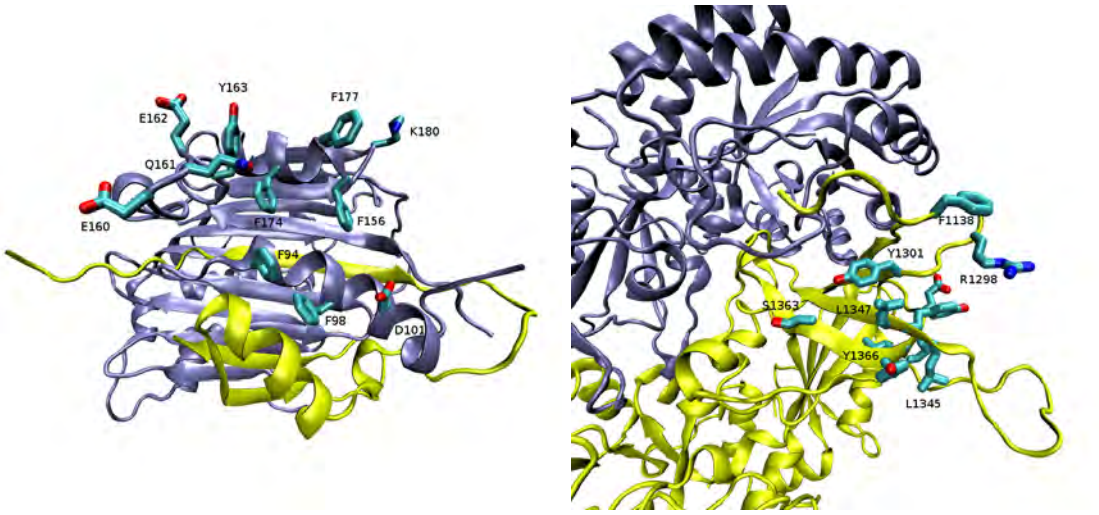


Figure 3.41: Region 1 and 2 of conserved contacting regions between AdoMetDC and ODC. See text for further details.

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.

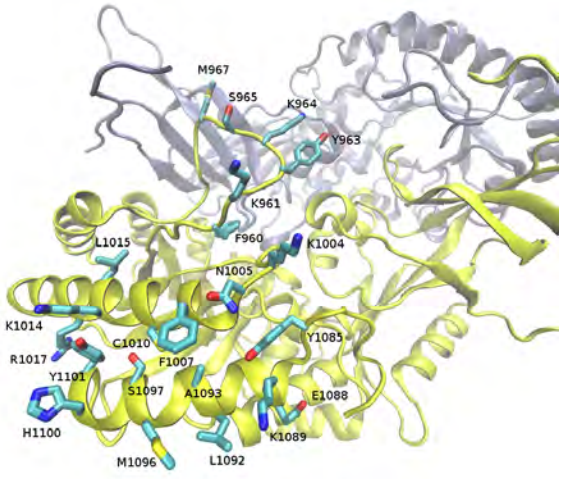


Figure 3.43: ODC patch of region 2. Numbering as for *P. falciparum*. For ODC only those conserved residues whose sidechains are oriented away from the surface are listed.

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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-

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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.

Bibliography

Ackermann, J. M., Pegg, A. E., and McCloskey, D. E. (2003) Drugs affecting the cell cycle via actions on the polyamine metabolic pathway. *Prog Cell Cycle Res*, **5**, 461–8. 4

Ahn, H. J., Kim, K. H., Lee, J., Ha, J. Y., Lee, H. H., Kim, D., Yoon, H. J., Kwon, A. R., and Suh, S. W. (2004) Crystal structure of agmatinase reveals structural conservation and inhibition mechanism of the ureohydrolase superfamily. *J Biol Chem*, **279** (48), 50505–13. 24

Albeck, S., Dym, O., Unger, T., Snapir, Z., Bercovich, Z., and Kahana, C. (2008) Crystallographic and biochemical studies revealing the structural basis for antizyme inhibitor function. *Protein Sci*, **17** (5), 793–802. 76

Almrud, J. J., Oliveira, M. A., Kern, A. D., Grishin, N. V., Phillips, M. A., and Hackert, M. L. (2000) Crystal structure of human ornithine decarboxylase at 2.1 a resolution: structural insights to antizyme binding. *J Mol Biol*, **295** (1), 7–16. 74, 86

Amarantos, I. and Kalpaxis, D. L. (2000) Photoaffinity polyamines: interactions with AcPhe-tRNA free in solution or bound at the p-site of *Escherichia coli* ribosomes. *Nucleic Acids Res*, **28** (19), 3733–3742. 3

Amarantos, I., Zarkadis, I. K., and Kalpaxis, D. L. (2002) The identification of spermine binding sites in 16S rRNA allows interpretation of the spermine effect on ribosomal 30S subunit functions. *Nucleic Acids Res*, **30** (13), 2832–2843. 3

Andrews, K. T., Tran, T. N., Lucke, A. J., Kahnberg, P., Le, G. T., Boyle, G. M., Gardiner, D. L., Skinner-Adams, T. S., and Fairlie, D. P. (2008) Potent antimalarial activity of histone deacetylase inhibitor analogues. *Antimicrob Agents Chemother*, **52** (4), 1454–1461. 21

Aravind, L., Iyer, L. M., Wellems, T. E., and Miller, L. H. (2003) *Plasmodium* biology: genomic gleanings. *Cell*, **115** (7), 771–85. 38, 39

Assaraf, Y. G., Golenser, J., Spira, D. T., Messer, G., and Bachrach, U. (1987) Cytostatic effect of DL-α-difluoromethylornithine against *Plasmodium falciparum* and its reversal by diamines and spermidine. *Parasitol Res*, **73** (4), 313–318. 7

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Bach, R., Canepa, C., and Glukhovtsev, M. (1999) Influence of electrostatic effects on activation barriers in enzymatic reactions: Pyridoxal 5́-phosphate-dependent decarboxylation of α-amino acids. *J Am Chem Soc*, **121** (28), 6542–6555. 75

Bachrach, U. (2005) Naturally occurring polyamines: interaction with macromolecules. *Curr Protein Pept Sci*, **6** (6), 559–566. 2

Baggio, R., Cox, J. D., Harper, S. L., Speicher, D. W., and Christianson, D. W. (1999) A new chromophoric assay for arginase activity. *Anal Biochem*, **276** (2), 251–253. 71

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, Jr, C. J., and Whetzel, P. (2003) PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res*, **31** (1), 212–5. 38, 85

Bailey, T. L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, **2**, 28–36. 19

Bale, S., Brooks, W., Hanes, J. W., Mahesan, A. M., Guida, W. C., and Ealick, S. E. (2009) Role of the sulfonium center in determining the ligand specificity of human *S*-adenosylmethionine decarboxylase. *Biochemistry*, **48** (27), 6423–6430. 80

Bale, S., Lopez, M. M., Makhatadze, G. I., Fang, Q., Pegg, A. E., and Ealick, S. E. (2008) Structural basis for putrescine activation of human *S*-adenosylmethionine decarboxylase. *Biochemistry*, **47** (50), 13404–13417. 81

Bastien, O., Aude, J.-C., Roy, S., and Maréchal, E. (2004a) Fundamentals of massive automatic pairwise alignments of protein sequences: theoretical significance of z-value statistics. *Bioinformatics*, **20** (4), 534–537. 20

Bastien, O., Lespinats, S., Roy, S., Métayer, K., Fertil, B., Codani, J.-J., and Maréchal, E. (2004b) Analysis of the compositional biases in *Plasmodium falciparum* genome and proteome using *Arabidopsis thaliana* as a reference. *Gene*, **336** (2), 163–173. 19

Bastien, O., Roy, S., and Maréchal, E. (2005) Construction of non-symmetric substitution matrices derived from proteomes with biased amino acid distributions. *C R Biol*, **328** (5), 445–453. 20

Bechet, J. and Wiame, J. M. (1965) Indication of a specific regulatory binding protein for ornithinetranscarbamylase in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun*, **21** (3), 226–234. 28

Bennett, E. M., Ekstrom, J. L., Pegg, A. E., and Ealick, S. E. (2002) Monomeric *S*-adenosylmethionine decarboxylase from plants provides an alternative to putrescine stimulation. *Biochemistry*, **41** (49), 14509–17. 78, 81, 82, 83, 85

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Beswick, T. C., Willert, E. K., and Phillips, M. A. (2006) Mechanisms of allosteric regulation of *Trypanosoma cruzi S*-adenosylmethionine decarboxylase. *Biochemistry*, **45** (25), 7797–7807. 82

Bewley, M. C., Jeffrey, P. D., Patchett, M. L., Kanyo, Z. F., and Baker, E. N. (1999) Crystal structures of *Bacillus caldovelox* arginase in complex with substrate and inhibitors reveal new insights into activation, inhibition and catalysis in the arginase superfamily. *Structure*, **7** (4), 435–48. 25, 28, 31, 40

Bewley, M. C., Lott, J. S., Baker, E. N., and Patchett, M. L. (1996) The cloning, expression and crystallisation of a thermostable arginase. *FEBS Lett*, **386** (2-3), 215–8. 25

Birkholtz, L., Joubert, F., Neitz, A. W., and Louw, A. I. (2003) Comparative properties of a three-dimensional model of *Plasmodium falciparum* ornithine decarboxylase. *Proteins*, **50** (3), 464–73. 4, 83, 85, 87, 88

Birkholtz, L. M., Wrenger, C., Joubert, F., Wells, G. A., Walter, R. D., and Louw, A. I. (2004) Parasite-specific inserts in the bifunctional *S*-adenosylmethionine decarboxylase/ornithine decarboxylase of *Plasmodium falciparum* modulate catalytic activities and domain interactions. *Biochem J*, **377** (Pt 2), 439–48. 4, 19, 39, 83, 85, 103, 105, 123, 124

Bitonti, A. J., Dumont, J. A., Bush, T. L., Edwards, M. L., Stemerick, D. M., McCann, P. P., and Sjoerdsma, A. (1989) Bis(benzyl)polyamine analogs inhibit the growth of chloroquine-resistant human malaria parasites (*Plasmodium falciparum*) *in vitro* and in combination with α-difluoromethylornithine cure murine malaria. *Proc Natl Acad Sci U S A*, **86** (2), 651–655. 8

Bitonti, A. J., McCann, P. P., and Sjoerdsma, A. (1987) *Plasmodium falciparum* and *Plasmodium berghei*: effects of ornithine decarboxylase inhibitors on erythrocytic schizogony. *Exp Parasitol*, **64** (2), 237–243. 7

Bjelic, S. and Aqvist, J. (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry*, **43** (46), 14521–14528. 21

Bogan, A. A. and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol*, **280** (1), 1–9. 124, 125

Böhm, M., rzebecher, J. S., and Klebe, G. (1999) Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J Med Chem*, **42** (3), 458–477. 9

Borkovich, K. A. and Weiss, R. L. (1987) Relationship between two major immunoreactive forms of arginase in *Neurospora crassa*. *J Bacteriol*, **169** (12), 5510–7. 30

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Bourne, P. and Weissig, H., editors (2003) In *Structural Bioinformatics*. Wiley. 11, 14

Callebaut, I., Prat, K., Meurice, E., Mornon, J.-P., and Tomavo, S. (2005) Prediction of the general transcription factors associated with rna polymerase ii in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. *BMC Genomics*, **6**, 100. 19

Cama, E., Colleluori, D. M., Emig, F. A., Shin, H., Kim, S. W., Kim, N. N., Traish, A. M., Ash, D. E., and Christianson, D. W. (2003a) Human arginase II: crystal structure and physiological role in male and female sexual arousal. *Biochemistry*, **42** (28), 8445–51. 25, 27, 30, 31, 40

Cama, E., Emig, F. A., Ash, D. E., and Christianson, D. W. (2003b) Structural and functional importance of first-shell metal ligands in the binuclear manganese cluster of arginase I. *Biochemistry*, **42** (25), 7748–58. 26, 28

Cama, E., Pethe, S., Boucher, J. L., Han, S., Emig, F. A., Ash, D. E., Viola, R. E., Mansuy, D., and Christianson, D. W. (2004) Inhibitor coordination interactions in the binuclear manganese cluster of arginase. *Biochemistry*, **43** (28), 8987–99. 27

Cama, E., Shin, H., and Christianson, D. W. (2003c) Design of amino acid sulfonamides as transition-state analogue inhibitors of arginase. *J Am Chem Soc*, **125** (43), 13052–7. 27

Canutescu, A. A., Shelenkov, A. A., and Dunbrack, Jr, R. L. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci*, **12** (9), 2001–14. 37

Carvajal, N., Martínez, J., de Oca, F. M., Rodríguez, J., and Fernández, M. (1978) Subunit interactions and immobilised dimers of human liver arginase. *Biochim Biophys Acta*, **527** (1), 1–7. 28

Carvajal, N., Martínez, J., and Fernández, M. (1977) Immobilised monomers of human liver arginase. *Biochim Biophys Acta*, **481** (1), 177–183. 28, 71

Carvajal, N., Olate, J., Salas, M., Uribe, E., Lopez, V., Herrera, P., and Cerpa, J. (1999a) Chemical modification and site-directed mutagenesis of human liver arginase: evidence that the imidazole group of histidine-141 is not involved in substrate binding. *Arch Biochem Biophys*, **371** (2), 202–206. 27

Carvajal, N., Rodríguez, J. P., and Fernández, M. (1982) Hybrid, immobilised dimers of human liver arginase. *Biochim Biophys Acta*, **701** (3), 405–407. 28

Carvajal, N., Salas, M., López, V., Uribe, E., Herrera, P., Cerpa, J., and Fuentes, M. (1999b) Manganese-dependent inhibition of human liver arginase by borate. *J Inorg Biochem*, **77** (3-4), 163–167. 27

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Carvajal, N., Venegas, A., Oestreicher, G., and Plaza, M. (1971) Effect of manganese on the quaternary structure of human liver arginase. *Biochim Biophys Acta*, **250** (2), 437–442. 27, 28, 71

Cavalli, R. C., Burke, C. J., Kawamoto, S., Soprano, D. R., and Ash, D. E. (1994) Mutagenesis of rat liver arginase expressed in *Escherichia coli*: role of conserved histidines. *Biochemistry*, **33** (35), 10652–7. 26, 27, 28, 71

Childs, A. C., Mehta, D. J., and Gerner, E. W. (2003) Polyamine-dependent gene expression. *Cell Mol Life Sci*, **60** (7), 1394–406. 2, 3

Christopherson, R. I., Cinquin, O., Shojaei, M., Kuehn, D., and Menz, R. I. (2004) Cloning and expression of malarial pyrimidine enzymes. *Nucleosides Nucleotides Nucleic Acids*, **23** (8-9), 1459–1465. 18

Clare, J. J., Belcourt, M., and Farabaugh, P. J. (1988) Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast ty1 transposon. *Proc Natl Acad Sci U S A*, **85** (18), 6816–6820. 3

Clyne, T., Kinch, L. N., and Phillips, M. A. (2002) Putrescine activation of *Trypanosoma cruzi S*-adenosylmethionine decarboxylase. *Biochemistry*, **41** (44), 13207–13216. 81, 82

Coffino, P. (2001) Antizyme, a mediator of ubiquitin-independent proteasomal degradation. *Biochimie*, **83** (3-4), 319–323. 76

Cohen, S. (1998) In *A Guide to the Polyamines*. Oxford University Press. 1

Coleman, C. S., Stanley, B. A., Viswanath, R., and Pegg, A. E. (1994) Rapid exchange of subunits of mammalian ornithine decarboxylase. *J Biol Chem*, **269** (5), 3155–3158. 76

Cox, J. D., Cama, E., Colleluori, D. M., Pethe, S., Boucher, J. L., Mansuy, D., Ash, D. E., and Christianson, D. W. (2001) Mechanistic and metabolic inferences from the binding of substrate analogues and products to arginase. *Biochemistry*, **40** (9), 2689–701. 27, 30

Da'Dara, A. A., Henkle-Dührsen, K., and Walter, R. D. (1996) A novel trans-spliced mrna from *Onchocerca volvulus* encodes a functional *S*-adenosylmethionine decarboxylase. *Biochem J*, **320 (Pt 2)**, 519–530. 81

Da'dara, A. A. and Walter, R. D. (1998) Molecular and biochemical characterization of *S*-adenosylmethionine decarboxylase from the free-living nematode *Caenorhabditis elegans*. *Biochem J*, **336 (Pt 3)**, 545–550. 81

D'Agostino, L., di Pietro, M., and Luccia, A. D. (2005) Nuclear aggregates of polyamines are supramolecular structures that play a crucial role in genomic dna protection and conformation. *FEBS J*, **272** (15), 3777–3787. 2

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

D'Agostino, L. and Luccia, A. D. (2002) Polyamines interact with dna as molecular aggregates. *Eur J Biochem*, **269** (17), 4317–4325. 2

Darwin, C. (1859) In *The Origin of Species by Means of Natural Selection: The Preservation of Favored Races in the Struggle for Life (Penguin Classics)*. Penguin Classics. 6

de Beer, T. A., Louw, A. I., and Joubert, F. (2006) Elucidation of sulfadoxine resistance with structural models of the bifunctional *Plasmodium falciparum* dihydropterin pyrophosphokinase-dihydropteroate synthase. *Bioorg Med Chem*, **14** (13), 4433–43. 19

de Beer, T. A. P., Wells, G. A., Burger, P. B., Joubert, F., Marechal, E., Birkholtz, L., and Louw, A. I. (2009) Antimalarial drug discovery: *in silico* structural biology and rational drug design. *Infect Disord Drug Targets*, **9** (3), 304–318. 19

de Terán, H. G., Nervall, M., Dunn, B. M., Clemente, J. C., and Aqvist, J. (2006a) Computational analysis of plasmepsin iv bound to an allophenylnorstatine inhibitor. *FEBS Lett*, **580** (25), 5910–5916. 21

de Terán, H. G., Nervall, M., Ersmark, K., Liu, P., Janka, L. K., Dunn, B., Hallberg, A., and Aqvist, J. (2006b) Inhibitor binding to the plasmepsin IV aspartic protease from *Plasmodium falciparum*. *Biochemistry*, **45** (35), 10529–10541. 21

Delano, W. (2002) The PyMOL Molecular Graphics System on World Wide Web http://www.pymol.org. 37

Delfino, R. T., Santos-Filho, O. A., and Figueroa-Villar, J. D. (2002) Molecular modeling of wild-type and antifolate resistant mutant *Plasmodium falciparum* DHFR. *Biophys Chem*, **98** (3), 287–300. 21

Desai, P. V., Patny, A., Gut, J., Rosenthal, P. J., Tekwani, B., Srivastava, A., and Avery, M. (2006) Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory. *J Med Chem*, **49** (5), 1576–1584. 21

Desai, P. V., Patny, A., Sabnis, Y., Tekwani, B., Gut, J., Rosenthal, P., Srivastava, A., and Avery, M. (2004) Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The ChemBridge database. *J Med Chem*, **47** (26), 6609–6615. 21

Di Costanzo, L., Moulin, M., Haertlein, M., Meilleur, F., and Christianson, D. W. (2007a) Expression, purification, assay, and crystal structure of perdeuterated human arginase I. *Arch Biochem Biophys*, **465** (1), 82–9. 27

Di Costanzo, L., Pique, M. E., and Christianson, D. W. (2007b) Crystal structure of human arginase I complexed with thiosemicarbazide reveals an unusual thiocarbonyl μ-sulfide ligand in the binuclear manganese cluster. *J Am Chem Soc*, **129** (20), 6388–9. 71

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Di Costanzo, L., Sabio, G., Mora, A., Rodriguez, P. C., Ochoa, A. C., Centeno, F., and Christianson, D. W. (2005) Crystal structure of human arginase I at 1.29-A resolution and exploration of inhibition in the immune response. *Proc Natl Acad Sci U S A*, **102** (37), 13058–63. 25, 26, 27, 35, 71

DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003) The price of innovation: new estimates of drug development costs. *J Health Econ*, **22** (2), 151–185. 9

Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., Lwin, K. M., Ariey, F., Hanpithakpong, W., Lee, S. J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S. S., Yeung, S., Singhasivanon, P., Day, N. P. J., Lindegardh, N., Socheat, D., and White, N. J. (2009) Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*, **361** (5), 455–467. 5

Dowling, D., Costanzo, L., Gennadios, H., and Christianson, D. (2008) Evolution of the arginase fold and functional diversity. *Cell Mol Life Sci*. 24

Dror, R. O., Jensen, M. O., and Shaw, D. E. (2009) Elucidating membrane protein function through long-timescale molecular dynamics simulation. *Conf Proc IEEE Eng Med Biol Soc*, **1**, 2340–2342. 14, 130

Dunathan, H. C. (1966) Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc Natl Acad Sci U S A*, **55** (4), 712–716. 75

Ekstrom, J. L., Mathews, I. I., Stanley, B. A., Pegg, A. E., and Ealick, S. E. (1999) The crystal structure of human *S*-adenosylmethionine decarboxylase at 2.25 a resolution reveals a novel fold. *Structure*, **7** (5), 583–595. 78, 81

Ekstrom, J. L., Tolbert, W. D., Xiong, H., Pegg, A. E., and Ealick, S. E. (2001) Structure of a human *S*-adenosylmethionine decarboxylase self-processing ester intermediate and mechanism of putrescine stimulation of processing as revealed by the H243A mutant. *Biochemistry*, **40** (32), 9495–504. 79

Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., and Brooks, C. L. (2004) Assessing scoring functions for protein-ligand interactions. *J Med Chem*, **47** (12), 3032–3047. 9

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340** (6230), 245–246. 15

Fletcher, S., Cummings, C. G., Rivas, K., Katt, W. P., Hornéy, C., Buckner, F. S., Chakrabarti, D., Sebti, S. M., Gelb, M. H., Voorhis, W. C. V., and Hamilton, A. D. (2008) Potent, *Plasmodium*-selective farnesyltransferase inhibitors that arrest the growth of malaria parasites: structure-activity relationships of ethylenediamine-analogue scaffolds and homology model validation. *J Med Chem*, **51** (17), 5176–5197. 22

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Fletcher, S. and Hamilton, A. D. (2006) Targeting protein-protein interactions by rational design: mimicry of protein surfaces. *J R Soc Interface*, **3** (7), 215–33. 70

Flick, K., Ahuja, S., Chene, A., Bejarano, M. T., and Chen, Q. (2004) Optimized expression of *Plasmodium falciparum* erythrocyte membrane protein 1 domains in *Escherichia coli*. *Malar J*, **3**, 50. 18

Fry, D. C. (2006) Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers*, **84** (6), 535–52. 70

Fukumoto, G. H. and Byus, C. V. (1996) A kinetic characterization of putrescine and spermidine uptake and export in human erythrocytes. *Biochim Biophys Acta*, **1282** (1), 48–56. 8

Fuller, J. C., Burgoyne, N. J., and Jackson, R. M. (2009) Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today*, **14** (3-4), 155–161. 18

Gabb, H. A., Jackson, R. M., and Sternberg, M. J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol*, **272** (1), 106–120. 17

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419** (6906), 498–511. 6, 83

Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D. C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G. G., Clayton, R., White, O., Smith, H. O., Fraser, C. M., Adams, M. D., Venter, J. C., and Hoffman, S. L. (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science*, **282** (5391), 1126–1132. 83

Giles, T. N. and Graham, D. E. (2008) Crenarchaeal arginine decarboxylase evolved from an *S*-adenosylmethionine decarboxylase enzyme. *J Biol Chem*, **283** (38), 25829–25838. 77

Gillet, J. M., Charlier, J., Boné, G., and Mulamba, P. L. (1983) *Plasmodium berghei*: inhibition of the sporogonous cycle by α-difluoromethylornithine. *Exp Parasitol*, **56** (2), 190–193. 7

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Glenn, M. P., Chang, S.-Y., Hornéy, C., Rivas, K., Yokoyama, K., Pusateri, E. E., Fletcher, S., Cummings, C. G., Buckner, F. S., Pendyala, P. R., Chakrabarti, D., Sebti, S. M., Gelb, M., Voorhis, W. C. V., and Hamilton, A. D. (2006) Structurally simple, potent, *Plasmodium* selective farnesyltransferase inhibitors that arrest the growth of malaria parasites. *J Med Chem*, **49** (19), 5710–5727. 22

Glenn, M. P., Chang, S.-Y., Hucke, O., Verlinde, C. L. M. J., Rivas, K., Hornéy, C., Yokoyama, K., Buckner, F. S., Pendyala, P. R., Chakrabarti, D., Gelb, M., Voorhis, W. C. V., Sebti, S. M., and Hamilton, A. D. (2005) Structurally simple farnesyltransferase inhibitors arrest the growth of malaria parasites. *Angew Chem Int Ed Engl*, **44** (31), 4903–4906. 22

Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996) Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit*, **9** (1), 1–5. 10

Gotoh, T., Sonoki, T., Nagasaki, A., Terada, K., Takiguchi, M., and Mori, M. (1996) Molecular cloning of cDNA for nonhepatic mitochondrial arginase (arginase II) and comparison of its induction with nitric oxide synthase in a murine macrophage-like cell line. *FEBS Lett*, **395** (2-3), 119–22. 30

Green, S. M., Eisenstein, E., McPhie, P., and Hensley, P. (1990) The purification and characterization of arginase from *Saccharomyces cerevisiae*. *J Biol Chem*, **265** (3), 1601–7. 28

Green, S. M., Ginsburg, A., Lewis, M. S., and Hensley, P. (1991) Roles of metal ions in the maintenance of the tertiary and quaternary structure of arginase from *Saccharomyces cerevisiae*. *J Biol Chem*, **266** (32), 21474–81. 28

Greenwood, B. M., Fidock, D. A., Kyle, D. E., Kappe, S. H., Alonso, P. L., Collins, F. H., and Duffy, P. E. (2008) Malaria: progress, perils, and prospects for eradication. *The Journal of clinical investigation*, **118** (4), 1266–1276. 4, 5, 6, 7

Grishin, N. V., Osterman, A. L., Brooks, H. B., Phillips, M. A., and Goldsmith, E. J. (1999) X-ray structure of ornithine decarboxylase from *Trypanosoma brucei*: the native structure and the structure in complex with α-difluoromethylornithine. *Biochemistry*, **38** (46), 15174–15184. 74, 75, 76

Grishin, N. V., Phillips, M. A., and Goldsmith, E. J. (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases. *Protein Sci*, **4** (7), 1291–1304. 73

Grody, W. W., Dizikes, G. J., and Cederbaum, S. D. (1987) Human arginase isozymes. *Isozymes Curr Top Biol Med Res*, **13**, 181–214. 30

Gupta, R. D., Krause-Ihle, T., Bergmann, B., Müller, I. B., Khomutov, A. R., Müller, S., Walter, R. D., and Lüersen, K. (2005) 3-aminooxy-1-aminopropane and derivatives have an antiproliferative effect on cultured *Plasmodium falciparum* by decreasing intracellular polyamine concentrations. *Antimicrob Agents Chemother*, **49** (7), 2857–2864. 7, 8

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Haider, N., Eschbach, M. L., Dias Sde, S., Gilberger, T. W., Walter, R. D., and Lüersen, K. (2005) The spermidine synthase of the malaria parasite *Plasmodium falciparum*: molecular and biochemical characterisation of the polyamine synthesis enzyme. *Mol Biochem Parasitol*, **142** (2), 224–36. 4

Hardman, J. G. and Limbird, L. E. (2001) In *Goodman and Gilman's the Pharmacological Basis of Therapeutics (McGraw-Hill International Editions)*. McGraw-Hill Publishing Co., 10th edition. 5

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., Kabaria, C. W., Manh, B. H., Elyazar, I. R., Brooker, S., Smith, D. L., Moyeed, R. A., and Snow, R. W. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS medicine*, **6** (3). 4, 5

Heby, O., Persson, L., and Rentala, M. (2007) Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of african sleeping sickness, Chagas´ disease, and leishmaniasis. *Amino Acids*, **33** (2), 359–366. 3, 4

Hecker, M., Nematollahi, H., Hey, C., Busse, R., and Racké, K. (1995) Inhibition of arginase by NG-hydroxy-L-arginine in alveolar macrophages: implications for the utilization of l-arginine for nitric oxide synthesis. *FEBS Lett*, **359** (2-3), 251–254. 30

Hoyt, M. A., Williams-Abbott, L. J., Pitkin, J. W., and Davis, R. H. (2000) Cloning and expression of the *S*-adenosylmethionine decarboxylase gene of *Neurospora crassa* and processing of its product. *Mol Gen Genet*, **263** (4), 664–673. 81

Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph*, **14** (1), 33–8, 27–8. 37

Hyde, J. E. (2007) Drug-resistant malaria - an insight. *FEBS J*, **274** (18), 4688–4698. 4, 5

Ivanov, I. and Klein, M. L. (2004) First principles computational study of the active site of arginase. *Proteins*, **54** (1), 1–7. 26

Jackson, K. L., Baldwin, J., Akella, R., Goldsmith, E. J., and Phillips, M. A. (2004) Multiple active site conformations revealed by distant site mutation in ornithine decarboxylase. *Biochemistry*, **43** (41), 12990–12999. 74, 76

Jackson, L. K., Brooks, H. B., Myers, D. P., and Phillips, M. A. (2003a) Ornithine decarboxylase promotes catalysis by binding the carboxylate in a buried pocket containing phenylalanine 397. *Biochemistry*, **42** (10), 2933–2940. 75

Jackson, L. K., Brooks, H. B., Osterman, A. L., Goldsmith, E. J., and Phillips, M. A. (2000) Altering the reaction specificity of eukaryotic ornithine decarboxylase. *Biochemistry*, **39** (37), 11247–11257. 75, 86

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Jackson, L. K., Goldsmith, E. J., and Phillips, M. A. (2003b) X-ray structure determination of *Trypanosoma brucei* ornithine decarboxylase bound to D-ornithine and to G418: insights into substrate binding and ODC conformational flexibility. *J Biol Chem*, **278** (24), 22037–22043. 76

Jackson, R. M., Gabb, H. A., and Sternberg, M. J. (1998) Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J Mol Biol*, **276** (1), 265–285. 89

Janne, J., Alhonen, L., Pietila, M., and Keinanen, T. A. (2004) Genetic approaches to the cellular functions of polyamines in mammals. *Eur J Biochem*, **271** (5), 877–94. 2

Jean, L., Withers-Martinez, C., Hackett, F., and Blackman, M. J. (2005) Unique insertions within *Plasmodium falciparum* subtilisin-like protease-1 are crucial for enzyme maturation and activity. *Mol Biochem Parasitol*, **144** (2), 187–97. 39

Jenkinson, C. P., Grody, W. W., and Cederbaum, S. D. (1996) Comparative properties of arginases. *Comp Biochem Physiol B Biochem Mol Biol*, **114** (1), 107–32. 27

Kameji, T. and Pegg, A. E. (1987) Effect of putrescine on the synthesis of *S*-adenosylmethionine decarboxylase. *Biochem J*, **243** (1), 285–288. 81

Kanyo, Z. F., Scolnick, L. R., Ash, D. E., and Christianson, D. W. (1996) Structure of a unique binuclear manganese cluster in arginase. *Nature*, **383** (6600), 554–7. Enter text here. 25, 28, 31, 40

Kasam, V., Salzemann, J., Botha, M., Dacosta, A., Degliesposti, G., Isea, R., Kim, D., Maass, A., Kenyon, C., Rastelli, G., Hofmann-Apitius, M., and Breton, V. (2009) WISDOM-II: screening against multiple targets implicated in malaria using computational grid infrastructures. *Malar J*, **8**, 88. 10

Kern, A. D., Oliveira, M. A., Coffino, P., and Hackert, M. L. (1999) Structure of mammalian ornithine decarboxylase at 1.6 a resolution: stereochemical implications of PLP-dependent amino acid decarboxylases. *Structure*, **7** (5), 567–581. 73, 75, 76, 125

Keskin, O., Gursoy, A., Ma, B., and Nussinov, R. (2007) Towards drugs targeting multiple proteins in a systems biology approach. *Curr Top Med Chem*, **7** (10), 943–51. 70

Kihara, D. and Skolnick, J. (2003) The PDB is a covering set of small protein structures. *J Mol Biol*, **334** (4), 793–802. 19

Kim, A. D., Graham, D. E., Seeholzer, S. H., and Markham, G. D. (2000) *S*-Adenosylmethionine decarboxylase from the archaeon *Methanococcus jannaschii*: identification of a novel family of pyruvoyl enzymes. *J Bacteriol*, **182** (23), 6667–6672. 81

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Kim, N. N., Cox, J. D., Baggio, R. F., Emig, F. A., Mistry, S. K., Harper, S. L., Speicher, D. W., Morris, Jr, S. M., Ash, D. E., Traish, A., and Christianson, D. W. (2001) Probing erectile function: *S*-(2-boronoethyl)-L-cysteine binds to arginase as a transition state analogue and enhances smooth muscle relaxation in human penile corpus cavernosum. *Biochemistry*, **40** (9), 2678–88. 71

Kinch, L. N. and Phillips, M. A. (2000) Single-turnover kinetic analysis of *Trypanosoma cruzi S*-adenosylmethionine decarboxylase. *Biochemistry*, **39** (12), 3336–3343. 81

Klebe, G. (2006) Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today*, **11** (13-14), 580–594. 9

Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol*, **19** (2), 120–127. 14, 130

Korsinczky, M., Fischer, K., Chen, N., Baker, J., Rieckmann, K., and Cheng, Q. (2004) Sulfadoxine resistance in *Plasmodium vivax* is associated with a specific amino acid in dihydropteroate synthase at the putative sulfadoxine-binding site. *Antimicrob Agents Chemother*, **48** (6), 2214–2222. 21

Kozbial, P. Z. and Mushegian, A. R. (2005) Natural history of *S*-adenosylmethionine-binding proteins. *BMC Struct Biol*, **5**, 19. 77

Krause, T., Luersen, K., Wrenger, C., Gilberger, T. W., Muller, S., and Walter, R. D. (2000) The ornithine decarboxylase domain of the bifunctional ornithine decarboxylase/*S*-adenosylmethionine decarboxylase of *Plasmodium falciparum*: recombinant expression and catalytic properties of two different constructs. *Biochem J*, **352 Pt 2**, 287–92. 4, 83

Lavulo, L. T., Sossong, Jr, T. M., Brigham-Burke, M. R., Doyle, M. L., Cox, J. D., Christianson, D. W., and Ash, D. E. (2001) Subunit-subunit interactions in trimeric arginase. Generation of active monomers by mutation of a single amino acid. *J Biol Chem*, **276** (17), 14242–8. 28, 58, 59, 71

Leach, A. (2001) In *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, 2nd edition. 11, 12

Lee, J., Michael, A. J., Martynowski, D., Goldsmith, E. J., and Phillips, M. A. (2007) Phylogenetic diversity and the structural basis of substrate specificity in the β/α-barrel fold basic amino acid decarboxylases. *J Biol Chem*, **282** (37), 27115–27125. 73, 74

Lemcke, T., Christensen, I. T., and Jørgensen, F. S. (1999) Towards an understanding of drug resistance in malaria: three-dimensional structure of *Plasmodium falciparum* dihydrofolate reductase by homology building. *Bioorg Med Chem*, **7** (6), 1003–1011. 21

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Lensink, M. F., Méndez, R., and Wodak, S. J. (2007) Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins*, **69** (4), 704–718. 15, 16, 18

Li, H., Meininger, C. J., Hawker, Jr, J. R., Haynes, T. E., Kepka-Lenhart, D., Mistry, S. K., Morris, Jr, S. M., and Wu, G. (2001) Regulatory role of arginase I and II in nitric oxide, polyamine, and proline syntheses in endothelial cells. *Am J Physiol Endocrinol Metab*, **280** (1), E75–82. 30

Li, R., Chen, X., Gong, B., Selzer, P. M., Li, Z., Davidson, E., Kurzban, G., Miller, R. E., Nuzum, E. O., McKerrow, J. H., Fletterick, R. J., Gillmor, S. A., Craik, C. S., Kuntz, I. D., Cohen, F. E., and Kenyon, G. L. (1996) Structure-based design of parasitic protease inhibitors. *Bioorg Med Chem*, **4** (9), 1421–1427. 21

Lijnzaad, P., Berendsen, H. J., and Argos, P. (1996) Hydrophobic patches on the surfaces of protein structures. *Proteins*, **25** (3), 389–397. 123

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **46** (1-3), 3–26. 9, 10

Liu, H., Elstner, M., Kaxiras, E., Frauenheim, T., Hermans, J., and Yang, W. (2001) Quantum mechanics simulation of protein dynamics on long timescale. *Proteins*, **44** (4), 484–489. 11

Livingstone (2000) The characterization of chemical structures using molecular properties. a survey. *J Chem Inf Comput Sci*, **40** (2), 195–209. 9

Lowa, P. M., Gillet, J., Boné, G., and Schechter, P. J. (1986) α-Difluoromethylornithine inhibits the first part of exoerythrocytic schizogony of *Plasmodium berghei* in rodents. *Ann Soc Belg Med Trop*, **66** (4), 301–308. 7

Luccia, A. D., Picariello, G., Iacomino, G., Formisano, A., Paduano, L., and D'Agostino, L. (2009) The *in vitro* nuclear aggregates of polyamines. *FEBS J*, **276** (8), 2324–2335. 2

Luksch, T., Chan, N.-S., Brass, S., Sotriffer, C. A., Klebe, G., and Diederich, W. E. (2008) Computer-aided design and synthesis of nonpeptidic plasmepsin II and IV inhibitors. *ChemMedChem*, **3** (9), 1323–1336. 21

Marathe, S., Yu, Y. G., Turner, G. E., Palmier, C., and Weiss, R. L. (1998) Multiple forms of arginase are differentially expressed from a single locus in *Neurospora crassa*. *J Biol Chem*, **273** (45), 29776–29785. 30

Mardis, K. L., Sutton, H. M., Zuo, X., Lindsey, J. S., and Tiede, D. M. (2009) Solution-state conformational ensemble of a hexameric porphyrin array characterized using molecular dynamics and X-ray scattering. *J Phys Chem A*, **113** (11), 2516–2523. 11, 128

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Markham, G. D., Tabor, C. W., and Tabor, H. (1982) *S*-adenosylmethionine decarboxylase of *Escherichia coli*. studies on the covalently linked pyruvate required for activity. *J Biol Chem*, **257** (20), 12063–12068. 81, 82

Marton, L. J. and Pegg, A. E. (1995) Polyamines as targets for therapeutic intervention. *Annu Rev Pharmacol Toxicol*, **35**, 55–91. 7

McKie, J. H., Douglas, K. T., Chan, C., Roser, S. A., Yates, R., Read, M., Hyde, J. E., Dascombe, M. J., Yuthavong, Y., and Sirawaraporn, W. (1998) Rational drug design approach for overcoming drug resistance: application to pyrimethamine resistance in malaria. *J Med Chem*, **41** (9), 1367–1370. 21

Mehlin, C., Boni, E., Buckner, F. S., Engel, L., Feist, T., Gelb, M. H., Haji, L., Kim, D., Liu, C., Mueller, N., Myler, P. J., Reddy, J. T., Sampson, J. N., Subramanian, E., Voorhis, W. C. V., Worthey, E., Zucker, F., and Hol, W. G. J. (2006) Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Mol Biochem Parasitol*, **148** (2), 144–160. 18

Melby, T. and Westby, M. (2009) Inhibitors of viral entry. *Handb Exp Pharmacol*, (189), 177–202. 18

Méndez, R., Leplae, R., Lensink, M. F., and Wodak, S. J. (2005) Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, **60** (2), 150–169. 16, 18

Méndez, R., Leplae, R., Maria, L. D., and Wodak, S. J. (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins*, **52** (1), 51–67. 16, 18

Messenguy, F. and Wiame, J. (1969) The control of ornithinetranscarbamylase activity by arginase in *Saccharomyces cerevisiae*. *FEBS Lett*, **3** (1), 47–49. 28

Mora, A., del Ara Rangel, M., Fuentes, J. M., Soler, G., and Centeno, F. (2000) Implications of the *S*-shaped domain in the quaternary structure of human arginase. *Biochim Biophys Acta*, **1476** (2), 181–90. 28, 40

Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007) Hot spots–a review of the protein-protein interface determinant amino-acid residues. *Proteins*, **68** (4), 803–812. 124

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. *Proteins*, **12** (4), 345–364. 88

Morris, S. M. (2002) Regulation of enzymes of the urea cycle and arginine metabolism. *Annu Rev Nutr*, **22**, 87–105. 30

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Müller, I. B., Gupta, R. D., Lüersen, K., Wrenger, C., and Walter, R. D. (2008) Assessing the polyamine metabolism of *Plasmodium falciparum* as chemotherapeutic target. *Mol Biochem Parasitol*, **160** (1), 1–7. 7, 8

Müller, I. B., Walter, R. D., and Wrenger, C. (2005) Structural metal dependency of the arginase from the human malaria parasite *Plasmodium falciparum*. *Biol Chem*, **386** (2), 117–26. 4, 29, 30, 37, 38, 48, 57, 59, 66, 67

Müller, S., Da'dara, A., Lüersen, K., Wrenger, C., Gupta, R. D., Madhubala, R., and Walter, R. D. (2000) In the human malaria parasite *Plasmodium falciparum*, polyamines are synthesized by a bifunctional ornithine decarboxylase, *S*-adenosylmethionine decarboxylase. *J Biol Chem*, **275** (11), 8097–8102. 4, 82

Müller, S., Liebau, E., Walter, R. D., and Krauth-Siegel, R. L. (2003) Thiol-based redox metabolism of protozoan parasites. *Trends Parasitol*, **19** (7), 320–8. 3, 4

Munder, M., Eichmann, K., Morán, J. M., Centeno, F., Soler, G., and Modolell, M. (1999) Th1/Th2-regulated expression of arginase isoforms in murine macrophages and dendritic cells. *J Immunol*, **163** (7), 3771–3777. 30

Myers, D. P., Jackson, L. K., Ipe, V. G., Murphy, G. E., and Phillips, M. A. (2001) Long-range interactions in the dimer interface of ornithine decarboxylase are important for enzyme function. *Biochemistry*, **40** (44), 13230–13236. 76

Olszewski, K. L., Morrisey, J. M., Wilinski, D., Burns, J. M., Vaidya, A. B., Rabinowitz, J. D., and Llinás, M. (2009) Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. *Cell Host Microbe*, **5** (2), 191–199. 71

Orry, A. J. W., Abagyan, R. A., and Cavasotto, C. N. (2006) Structure-based development of target-specific compound libraries. *Drug Discov Today*, **11** (5-6), 261–266. 9

Osterman, A., Grishin, N. V., Kinch, L. N., and Phillips, M. A. (1994) Formation of functional cross-species heterodimers of ornithine decarboxylase. *Biochemistry*, **33** (46), 13662–13667. 76

Osterman, A. L., Brooks, H. B., Jackson, L., Abbott, J. J., and Phillips, M. A. (1999) Lysine-69 plays a key role in catalysis by ornithine decarboxylase through acceleration of the schiff base formation, decarboxylation, and product release steps. *Biochemistry*, **38** (36), 11814–11826. 74

Ouzounis, C. A., Kunin, V., Darzentas, N., and Goldovsky, L. (2006) A minimal estimate for the gene content of the last universal common ancestor–exobiology from a terrestrial perspective. *Res Microbiol*, **157** (1), 57–68. 24

Ouzounis, C. A. and Kyrpides, N. C. (1994) On the evolution of arginases and related enzymes. *J Mol Evol*, **39** (1), 101–4. 24

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Pankaskie, M. and Abdel-Monem, M. M. (1980) Inhibitors of polyamine biosynthesis. 8. irreversible inhibition of mammalian *S*-adenosyl-L-methionine decarboxylase by substrate analogues. *J Med Chem*, **23** (2), 121–127. 80

Park, M. H., Cooper, H. L., and Folk, J. E. (1981) Identification of hypusine, an unusual amino acid, in a protein from human lymphocytes and of spermidine as its biosynthetic precursor. *Proc Natl Acad Sci U S A*, **78** (5), 2869–2873. 3

Park, M. H., Cooper, H. L., and Folk, J. E. (1982) The biosynthesis of protein-bound hypusine (N_ϵ-(4-amino-2-hydroxybutyl)lysine). lysine as the amino acid precursor and the intermediate role of deoxyhypusine (n epsilon -(4-aminobutyl)lysine). *J Biol Chem*, **257** (12), 7217–7222. 3

Park, M. H., Wolff, E. C., and Folk, J. E. (1993) Is hypusine essential for eukaryotic cell proliferation? *Trends Biochem Sci*, **18** (12), 475–9. 3

Pegg, A. E. and Jacobs, G. (1983) Comparison of inhibitors of *S*-adenosylmethionine decarboxylase from different species. *Biochem J*, **213** (2), 495–502. 80

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, **26** (16), 1781–802. 47

Pizzi, E. and Frontali, C. (2000) Divergence of noncoding sequences and of insertions encoding nonglobular domains at a genomic region well conserved in plasmodia. *J Mol Evol*, **50** (5), 474–480. 83

Pizzi, E. and Frontali, C. (2001) Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res*, **11** (2), 218–29. 39, 83

Punta, M. and Rost, B. (2005) PROFcon: novel prediction of long-range contacts. *Bioinformatics*, **21** (13), 2960–2968. 116

R Development Core Team (2009) In *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 102

Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*, **261** (3), 470–489. 10

Rastelli, G., Sirawaraporn, W., Sompornpisut, P., Vilaivan, T., Kamchonwongpaisan, S., Quarrell, R., Lowe, G., Thebtaranonth, Y., and Yuthavong, Y. (2000) Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with *Plasmodium falciparum* dihydrofolate reductase: structural basis of antifolate resistance. *Bioorg Med Chem*, **8** (5), 1117–1128. 21, 22

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Rattanachuen, W., Jönsson, M., Swedberg, G., and Sirawaraporn, W. (2009) Probing the roles of non-homologous insertions in the n-terminal domain of *Plasmodium falciparum* hydroxymethylpterin pyrophosphokinase-dihydropteroate synthase. *Mol Biochem Parasitol*, **168** (2), 135–142. 39

Recsei, P. A. and Snell, E. E. (1984) Pyruvoyl enzymes. *Annu Rev Biochem*, **53**, 357–387. 79

Rom, E. and Kahana, C. (1994) Polyamines regulate the expression of ornithine decarboxylase antizyme *in vitro* by inducing ribosomal frame-shifting. *Proc Natl Acad Sci U S A*, **91** (9), 3959–3963. 3

Rowe, A. K., Rowe, S. Y., Snow, R. W., Korenromp, E. L., Schellenberg, J. R. A. R., Stein, C., Nahlen, B. L., Bryce, J., Black, R. E., and Steketee, R. W. (2006) The burden of malaria mortality among african children in the year 2000. *International journal of epidemiology*, **35** (3), 691–704. 4

Sabio, G., Mora, A., Rangel, M. A., Quesada, A., Marcos, C. F., Alonso, J. C., Soler, G., and Centeno, F. (2001) Glu-256 is a main structural determinant for oligomerisation of human arginase I. *FEBS Lett*, **501** (2-3), 161–165. 28, 39, 71

Sandmeier, E., Hale, T. I., and Christen, P. (1994) Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases. *Eur J Biochem*, **221** (3), 997–1002. 73

Santos-Filho, O. A., de Alencastro, R. B., and Figueroa-Villar, J. D. (2001) Homology modeling of wild type and pyrimethamine/cycloguanil-cross resistant mutant type *Plasmodium falciparum* dihydrofolate reductase. A model for antimalarial chemotherapy resistance. *Biophys Chem*, **91** (3), 305–317. 21

Schafer, B., Hauber, I., Bunk, A., Heukeshoven, J., Dusedau, A., Bevec, D., and Hauber, J. (2006) Inhibition of multidrug-resistant HIV-1 by interference with cellular *S*-adenosylmethionine decarboxylase activity. *J Infect Dis*, **194** (6), 740–50. 4

Schlick, T. (2002) In *Molecular Modeling and Simulation*. Springer, 1st edition. 11, 12, 14

Scolnick, L. R., Kanyo, Z. F., Cavalli, R. C., Ash, D. E., and Christianson, D. W. (1997) Altering the binuclear manganese cluster of arginase diminishes thermostability and catalytic function. *Biochemistry*, **36** (34), 10558–65. 25, 26, 28

Seiler, N. (2003a) Thirty years of polyamine-related approaches to cancer therapy. Retrospect and prospect. Part 1. Selective enzyme inhibitors. *Curr Drug Targets*, **4** (7), 537–564. 7

Seiler, N. (2003b) Thirty years of polyamine-related approaches to cancer therapy. Retrospect and prospect. Part 2. Structural analogues and derivatives. *Curr Drug Targets*, **4** (7), 565–585. 7

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Sekowska, A., Coppée, J. Y., Caer, J. P. L., Martin-Verstraete, I., and Danchin, A. (2000) *S*-adenosylmethionine decarboxylase of *Bacillus subtilis* is closely related to archaebacterial counterparts. *Mol Microbiol*, **36** (5), 1135–1147. 24, 81

Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*, **310** (1), 243–57. 31

Shoemaker, B. A. and Panchenko, A. R. (2007) Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol*, **3** (3), e42. 15

Singh, B., Sung, L. K., Matusop, A., Radhakrishnan, A., Shamsul, S. S. G., Cox-Singh, J., Thomas, A., and Conway, D. J. (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet*, **363** (9414), 1017–1024. 4, 21

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., and Hay, S. I. (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, **434** (7030), 214–7. 4

Snowden, M. and Green, D. V. (2008) The impact of diversity-based, high-throughput screening on drug discovery: "chance favours the prepared mind". *Curr Opin Drug Discov Devel*, **11** (4), 553–558. 9

Spector, E. B., Rice, S. C., Kern, R. M., Hendrickson, R., and Cederbaum, S. D. (1985) Comparison of arginase activity in red blood cells of lower mammals, primates, and man: evolution to high activity in primates. *Am J Hum Genet*, **37** (6), 1138–45. 30

Spector, E. B., Rice, S. C., Moedjono, S., Bernard, B., and Cederbaum, S. D. (1982) Biochemical properties of arginase in human adult and fetal tissues. *Biochem Med*, **28** (2), 165–75. 30

Stanley, B. A. and Pegg, A. E. (1991) Amino acid residues necessary for putrescine stimulation of human *S*-adenosylmethionine decarboxylase proenzyme processing and catalytic activity. *J Biol Chem*, **266** (28), 18502–18506. 79

Stanley, B. A., Shantz, L. M., and Pegg, A. E. (1994) Expression of mammalian *S*-adenosylmethionine decarboxylase in *Escherichia coli*. determination of sites for putrescine activation of activity and processing. *J Biol Chem*, **269** (11), 7901–7907. 81

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122** (6), 957–968. 15

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Stemmler, T. L., Sossong, T. M., Goldstein, J. I., Ash, D. E., Elgren, T. E., Kurtz, D. M., and Penner-Hahn, J. E. (1997) Exafs comparison of the dimanganese core structures of manganese catalase, arginase, and manganese-substituted ribonucleotide reductase and hemerythrin. *Biochemistry*, **36** (32), 9847–9858. 25

Sugiyama, T., Suzue, K., Okamoto, M., Inselburg, J., Tai, K., and Horii, T. (1996) Production of recombinant sera proteins of *Plasmodium falciparum* in *Escherichia coli* by using synthetic genes. *Vaccine*, **14** (11), 1069–1076. 18

Tabor, C. W. and Tabor, H. (1984) Polyamines. *Annu Rev Biochem*, **53**, 749–790. 1, 3

Tabor, C. W. and Tabor, H. (1985) Polyamines in microorganisms. *Microbiol Rev*, **49** (1), 81–99. 1

Takatsuka, Y., Yamaguchi, Y., Ono, M., and Kamio, Y. (2000) Gene cloning and molecular characterization of lysine decarboxylase from *Selenomonas ruminantium* delineate its evolutionary relationship to ornithine decarboxylases from eukaryotes. *J Bacteriol*, **182** (23), 6732–6741. 73

Theobald, D. L. and Wuttke, D. S. (2006a) Empirical bayes hierarchical models for regularizing maximum likelihood estimation in the matrix gaussian procrustes problem. *Proc Natl Acad Sci U S A*, **103** (49), 18521–18527. 89

Theobald, D. L. and Wuttke, D. S. (2006b) Theseus: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22** (17), 2171–2172. 89

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.q. *Nucleic Acids Res*, **25** (24), 4876–4882. 32, 87

Thorn, K. S. and Bogan, A. A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17** (3), 284–285. 124

Tiede, D., Mardis, K., and Zuo, X. (2009) X-ray scattering combined with coordinate-based analyses for applications in natural and artificial photosynthesis. *Photosynth Res*. 11, 128

Tolbert, W. D., Ekstrom, J. L., Mathews, I. I., Secrist, 3rd, J. A., Kapoor, P., Pegg, A. E., and Ealick, S. E. (2001) The structural basis for substrate specificity and inhibition of human *S*-adenosylmethionine decarboxylase. *Biochemistry*, **40** (32), 9484–94. 80, 81, 85

Tolbert, W. D., Zhang, Y., Cottet, S. E., Bennett, E. M., Ekstrom, J. L., Pegg, A. E., and Ealick, S. E. (2003) Mechanism of human *S*-adenosylmethionine decarboxylase proenzyme processing as revealed by the structure of the S68A mutant. *Biochemistry*, **42** (8), 2386–95. 79, 80

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Toms, A. V., Kinsland, C., McCloskey, D. E., Pegg, A. E., and Ealick, S. E. (2004) Evolutionary links as revealed by the structure of *Thermotoga maritima S*-adenosylmethionine decarboxylase. *J Biol Chem*, **279** (32), 33837–33846. 78

Toney, M. D. (2001) Computational studies on nonenzymatic and enzymatic pyridoxal phosphate catalyzed decarboxylations of 2-aminoisobutyrate. *Biochemistry*, **40** (5), 1378–1384. 75

Toney, M. D. (2005) Reaction specificity in pyridoxal phosphate enzymes. *Arch Biochem Biophys*, **433** (1), 279–287. 74

Toyoda, T., Brobey, R. K., Sano, G., Horii, T., Tomioka, N., and Itai, A. (1997) Lead discovery of inhibitors of the dihydrofolate reductase domain of *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase. *Biochem Biophys Res Commun*, **235** (3), 515–519. 21

Turner, G. E. and Weiss, R. L. (2006) Developmental expression of two forms of arginase in *Neurospora crassa*. *Biochim Biophys Acta*, **1760** (6), 848–857. 30

van Brummelen, A. C., Olszewski, K. L., Wilinski, D., Llinás, M., Louw, A. I., and Birkholtz, L.-M. (2009) Co-inhibition of *Plasmodium falciparum S*-adenosylmethionine decarboxylase/ornithine decarboxylase reveals perturbation-specific compensatory mechanisms by transcriptome, proteome, and metabolome analyses. *J Biol Chem*, **284** (7), 4635–4646. 7

van Poelje, P. D. and Snell, E. E. (1990) Pyruvoyl-dependent enzymes. *Annu Rev Biochem*, **59**, 29–59. 79

Vasilescu, J. and Figeys, D. (2006) Mapping protein-protein interactions by mass spectrometry. *Curr Opin Biotechnol*, **17** (4), 394–399. 15

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G. A., Gao, M., Hills, T., Brokx, S., Qiu, W., Sharma, S., Diassiti, A., Alam, Z., Melone, M., Mulichak, A., Wernimont, A., Bray, J., Loppnau, P., Plotnikova, O., Newberry, K., Sundararajan, E., Houston, S., Walker, J., Tempel, W., Bochkarev, A., Kozieradzki, I., Edwards, A., Arrowsmith, C., Roos, D., Kain, K., and Hui, R. (2007) Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Mol Biochem Parasitol*, **151** (1), 100–110. 19

Vernon, J. A., Golec, J. H., and Dimasi, J. A. (2009) Drug development costs when financial risk is measured using the fama-french three-factor model. *Health Econ*. 9

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph*, **8** (1), 52–6, 29. 88

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Wallace, H. M. (2007) Targeting polyamine metabolism: a viable therapeutic/preventative solution for cancer? *Expert Opin Pharmacother*, **8** (13), 2109–16. 4

Wallace, H. M., Fraser, A. V., and Hughes, A. (2003) A perspective of polyamine metabolism. *Biochem J*, **376** (Pt 1), 1–14. 2, 4

Wang, C. C. (1995) Molecular mechanisms and therapeutic approaches to the treatment of African trypanosomiasis. *Annu Rev Pharmacol Toxicol*, **35**, 93–127. 7

Weinberg, J. B., Lopansri, B. K., Mwaikambo, E., and Granger, D. L. (2008) Arginine, nitric oxide, carbon monoxide, and endothelial function in severe malaria. *Curr Opin Infect Dis*, **21** (5), 468–475. 71

Wells, G. A., Birkholtz, L. M., Joubert, F., Walter, R. D., and Louw, A. I. (2006) Novel properties of malarial *S*-adenosylmethionine decarboxylase as revealed by structural modelling. *J Mol Graph Model*, **24** (4), 307–18. 19, 83, 85, 86

Wells, G. A., Müller, I. B., Wrenger, C., and Louw, A. I. (2009) The activity of *Plasmodium falciparum* arginase is mediated by a novel inter-monomer salt-bridge between Glu295-Arg404. *FEBS J*, **276** (13), 3517–3530. 23

Willert, E. K., Fitzpatrick, R., and Phillips, M. A. (2007) Allosteric regulation of an essential trypanosome polyamine biosynthetic enzyme by a catalytically dead homolog. *Proc Natl Acad Sci U S A*, **104** (20), 8275–8280. 82

Willert, E. K. and Phillips, M. A. (2009) Cross-species activation of trypanosome *S*-adenosylmethionine decarboxylase by the regulatory subunit prozyme. *Mol Biochem Parasitol*, **168** (1), 1–6. 82

Withers-Martinez, C., Carpenter, E. P., Hackett, F., Ely, B., Sajid, M., Grainger, M., and Blackman, M. J. (1999) PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng*, **12** (12), 1113–1120. 18

Wrenger, C., Lüersen, K., Krause, T., Muller, S., and Walter, R. D. (2001) The *Plasmodium falciparum* bifunctional ornithine decarboxylase, *S*-adenosyl-L-methionine decarboxylase, enables a well balanced polyamine synthesis without domain-domain interaction. *J Biol Chem*, **276** (32), 29651–6. 4, 83

Wright, P. S., Byers, T. L., Cross-Doersen, D. E., McCann, P. P., and Bitonti, A. J. (1991) Irreversible inhibition of *S*-adenosylmethionine decarboxylase in *Plasmodium falciparum*-infected erythrocytes: growth inhibition *in vitro*. *Biochem Pharmacol*, **41** (11), 1713–1718. 7

Xiong, H. and Pegg, A. E. (1999) Mechanistic studies of the processing of human *S*-adenosylmethionine decarboxylase proenzyme. isolation of an ester intermediate. *J Biol Chem*, **274** (49), 35059–35066. 79

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Xiong, H., Stanley, B. A., Tekwani, B. L., and Pegg, A. E. (1997) Processing of mammalian and plant *S*-adenosylmethionine decarboxylase proenzymes. *J Biol Chem*, **272** (45), 28342–28348. 81

Xu, D., Tsai, C. J., and Nussinov, R. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng*, **10** (9), 999–1012. 125

Xue, H. Y. and Forsdyke, D. R. (2003) Low-complexity segments in *Plasmodium falciparum* proteins are primarily nucleic acid level adaptations. *Mol Biochem Parasitol*, **128** (1), 21–32. 39

Yadava, A. and Ockenhouse, C. F. (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infect Immun*, **71** (9), 4961–4969. 18

Yan, C., Wu, F., Jernigan, R. L., Dobbs, D., and Honavar, V. (2008) Characterization of protein-protein interfaces. *Protein J*, **27** (1), 59–70. 95

Young, L., Jernigan, R. L., and Covell, D. G. (1994) A role for surface hydrophobicity in protein-protein recognition. *Protein Sci*, **3** (5), 717–729. 123

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M. D., and Yuthavong, Y. (2003) Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol*, **10** (5), 357–65. 21, 39

Zaheer-ul, H., Uddin, R., Yuan, H., Petukhov, P. A., Choudhary, M. I., and Madura, J. D. (2008) Receptor-based modeling and 3D-QSAR for a quantitative production of the butyrylcholinesterase inhibitors based on genetic algorithm. *J Chem Inf Model*, **48** (5), 1092–1103. 10

Appendix A

Supplementary data for Chapter 2

A.1 Inter-monomer interactions in arginase

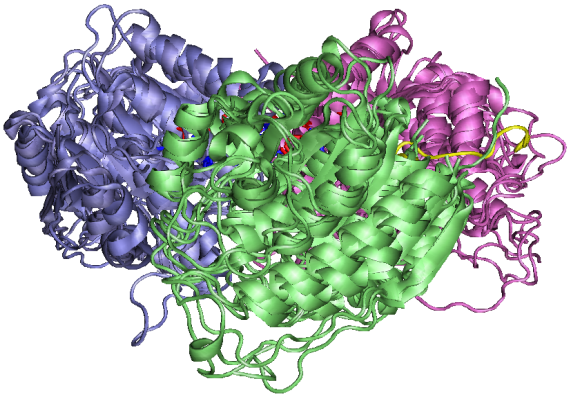


Figure A.1: A summary of inter-monomer interactions in arginase. Click on the image to active 3D. Sub-structures relevant to this work are highlighted in the model tree.

A.2 Co-ordination geometry of Mg^{2+}

A.2.1 Glu 295 Ala

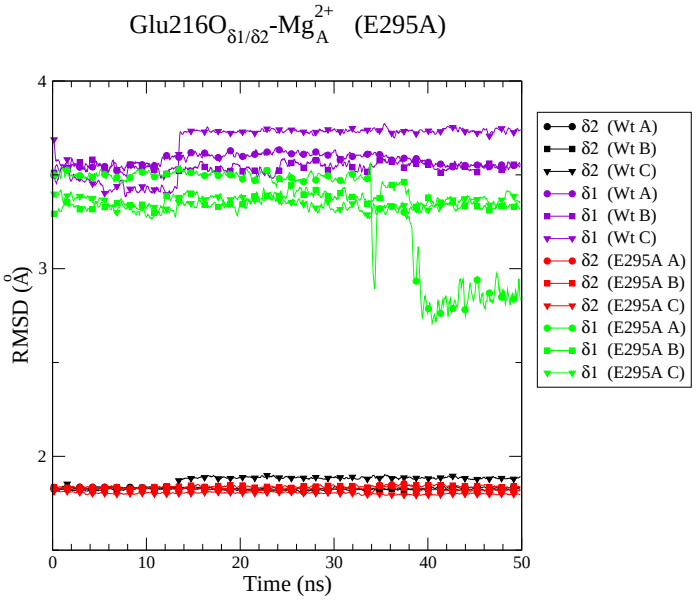


Figure A.2: Interaction between Asp 216 O and Mg_A^{2+} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

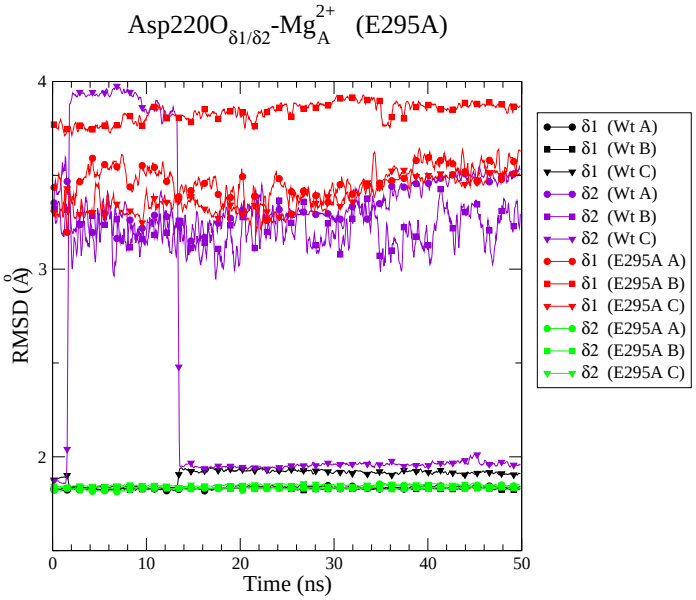


Figure A.3: Interaction between Asp 220 O and Mg_A^{2+} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

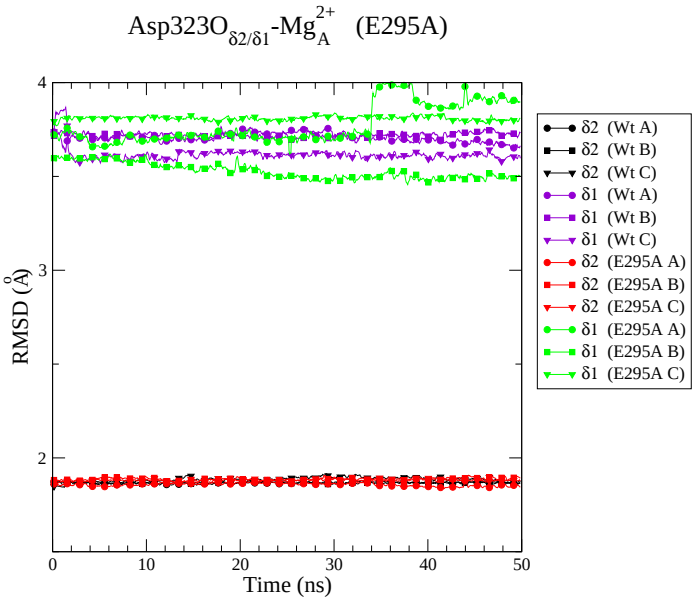


Figure A.4: Interaction between Asp 323 O and Mg^{2+}_{A} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

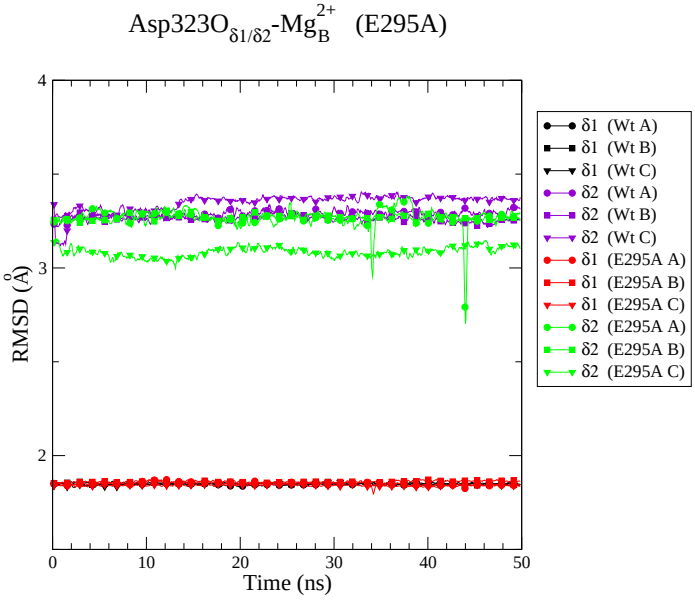


Figure A.5: Interaction between Asp 323 O and Mg^{2+}_{B} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

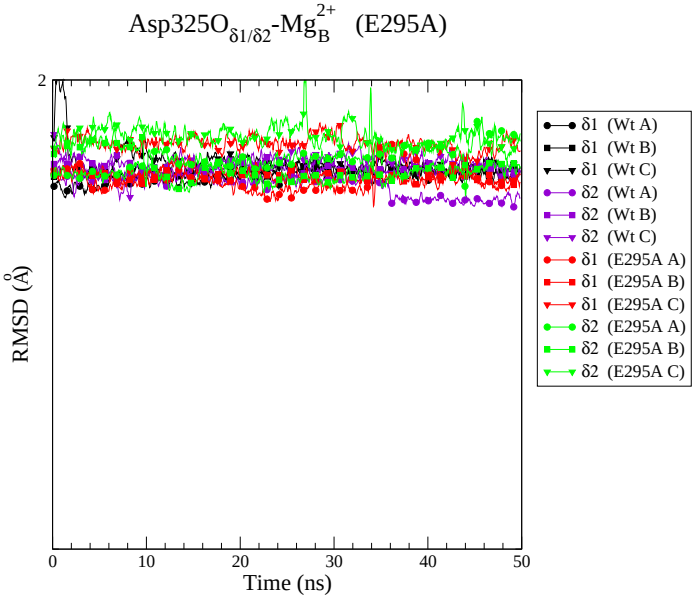


Figure A.6: Interaction between Asp 325 O and Mg^{2+}_{B} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

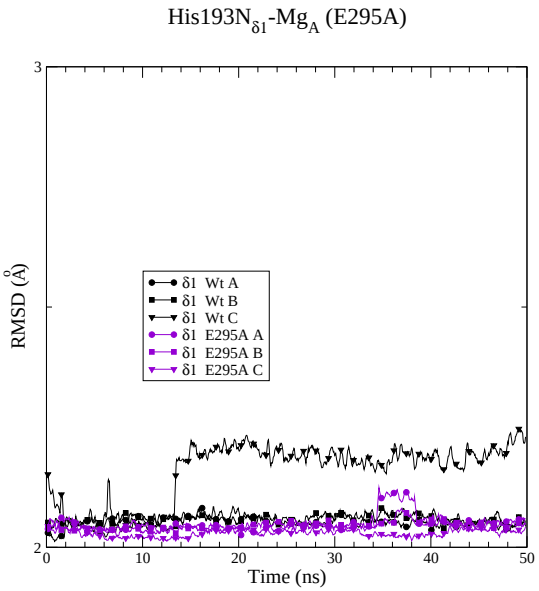


Figure A.7: Interaction between His 193 $N_{\delta1}$ and Mg_A^{2+} in *pf* Arg Glu 295 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

A.2.2 Glu 295 Ala/Arg 404 Ala

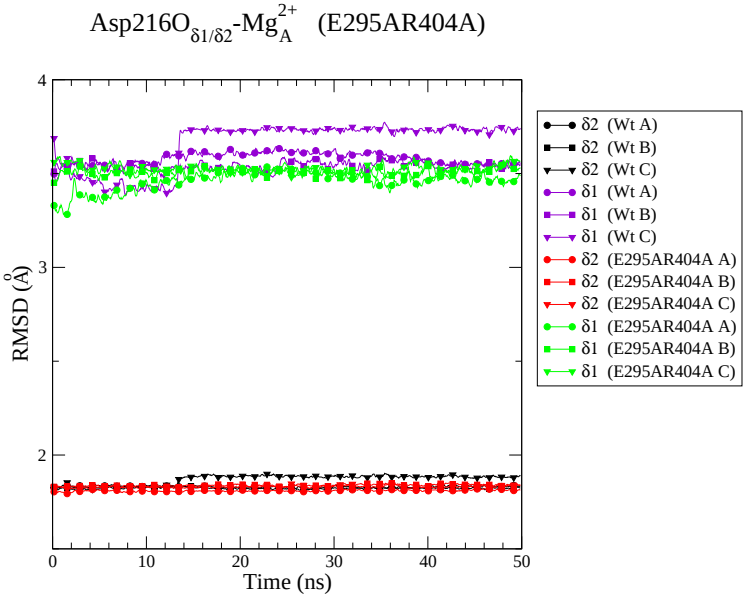


Figure A.8: Interaction between Asp 216 O and Mg_A^{2+} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

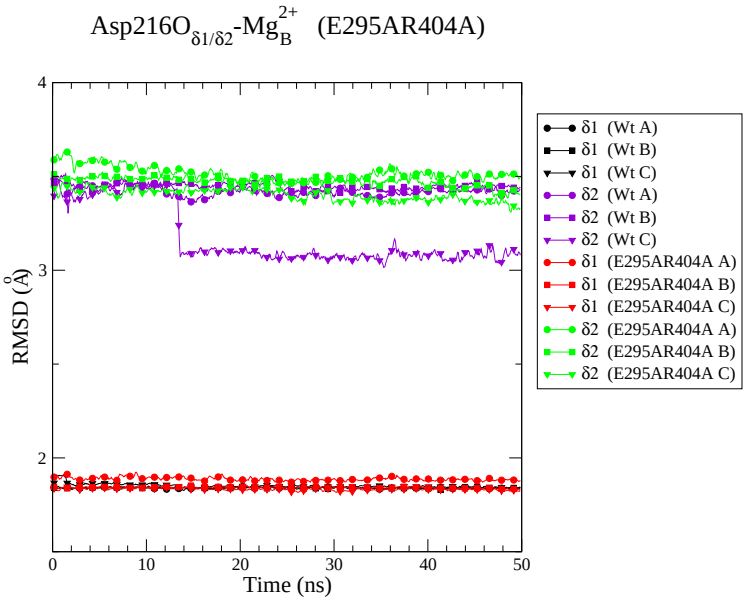


Figure A.9: Interaction between Asp 216 O and Mg^{2+}_{B} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

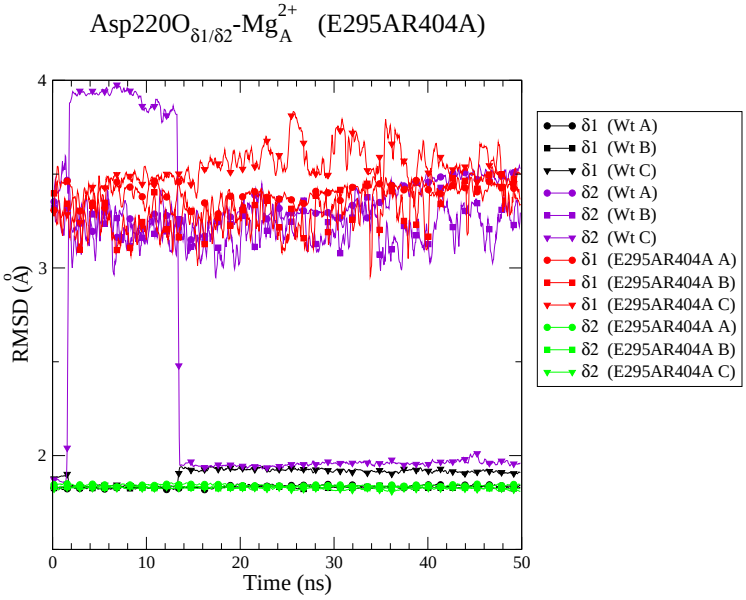


Figure A.10: Interaction between Asp 220 O and Mg^{2+}_{A} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

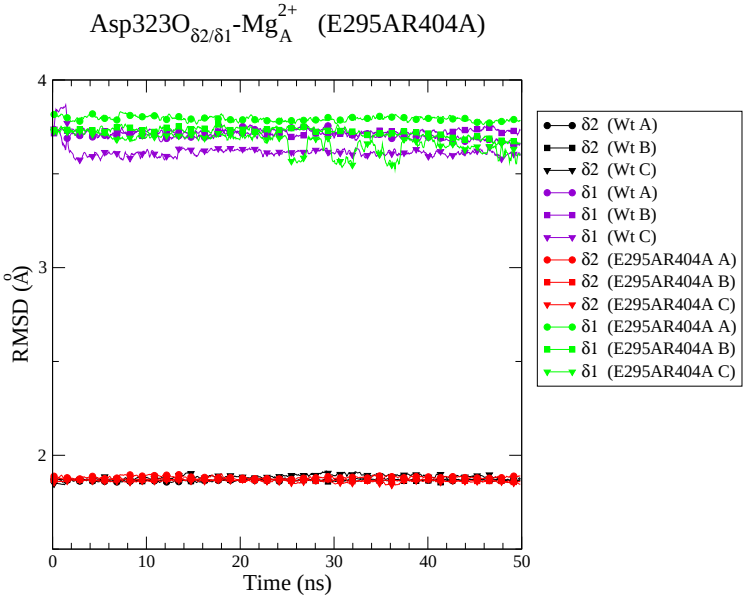


Figure A.11: Interaction between Asp 323 O and Mg^{2+}_{A} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

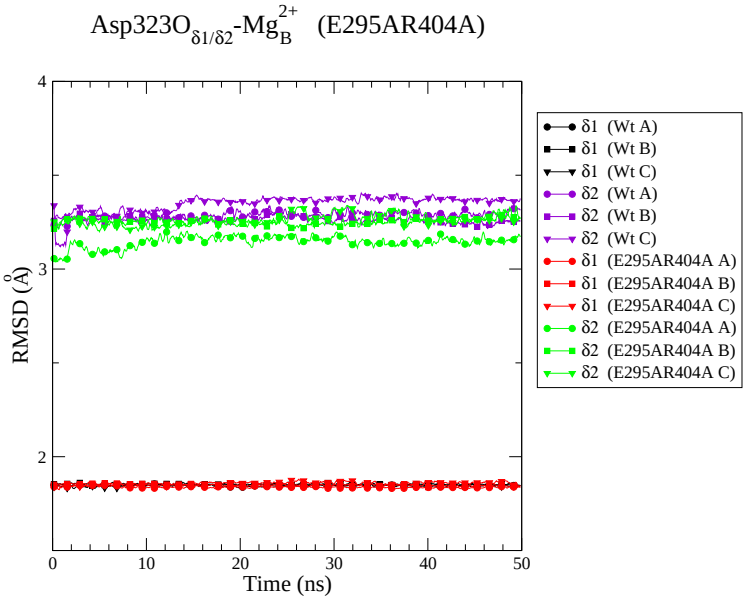


Figure A.12: Interaction between Asp 323 O and Mg_B^{2+} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

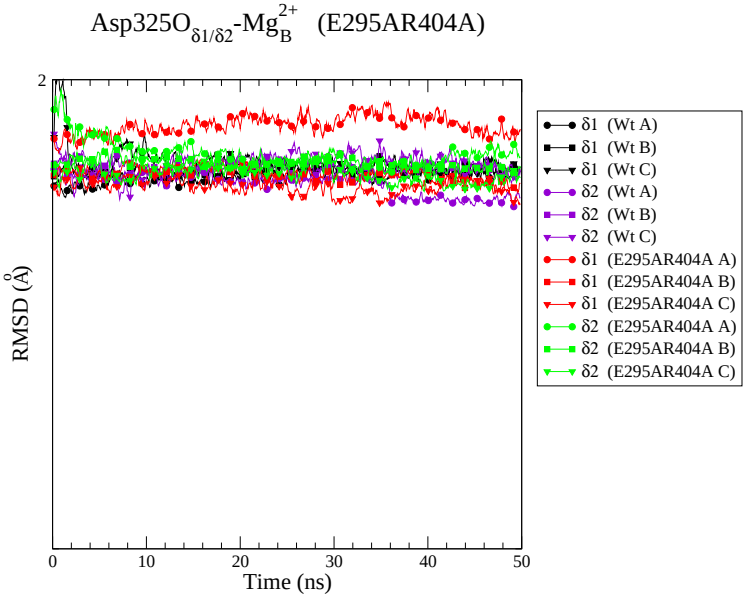


Figure A.13: Interaction between Asp 325 O and Mg_B^{2+} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

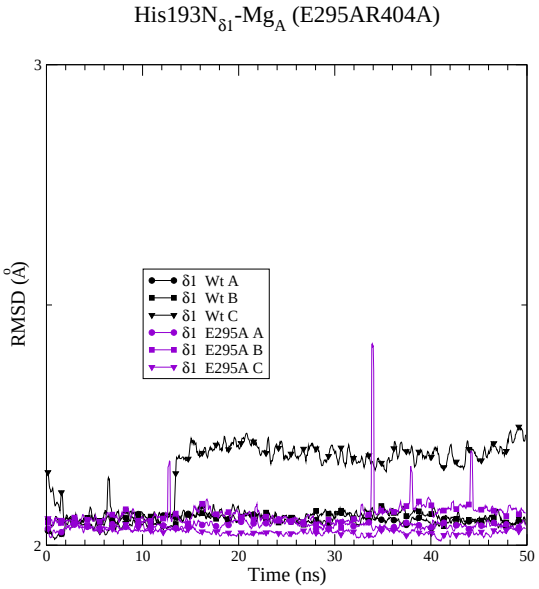


Figure A.14: Interaction between His 193 $N_{\delta1}$ and Mg_A^{2+} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

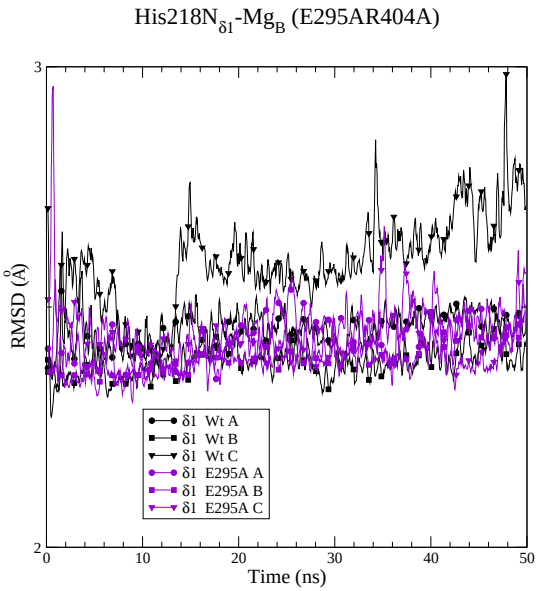


Figure A.15: Interaction between His 218 $N_{\delta 1}$ and Mg_B^{2+} in *pf* Arg Glu 295 Ala/Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta 1/\delta 2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

A.2.3 Glu 295 Arg

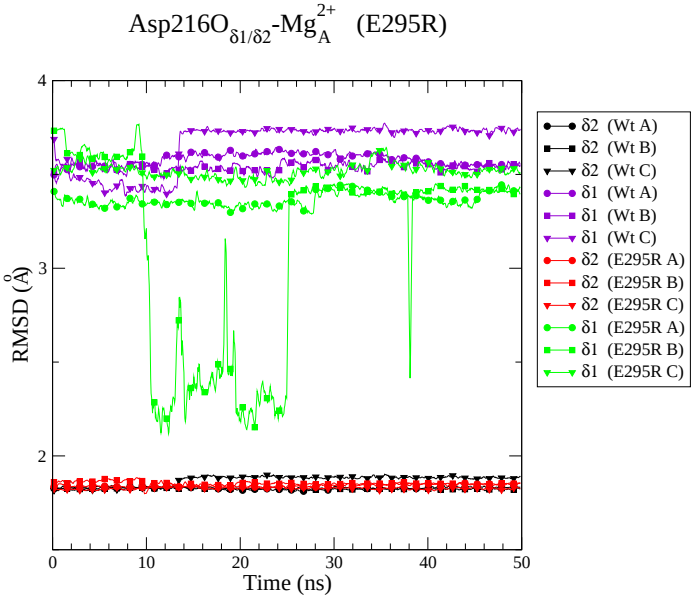


Figure A.16: Interaction between Asp 216 O and Mg_A^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta 1/\delta 2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

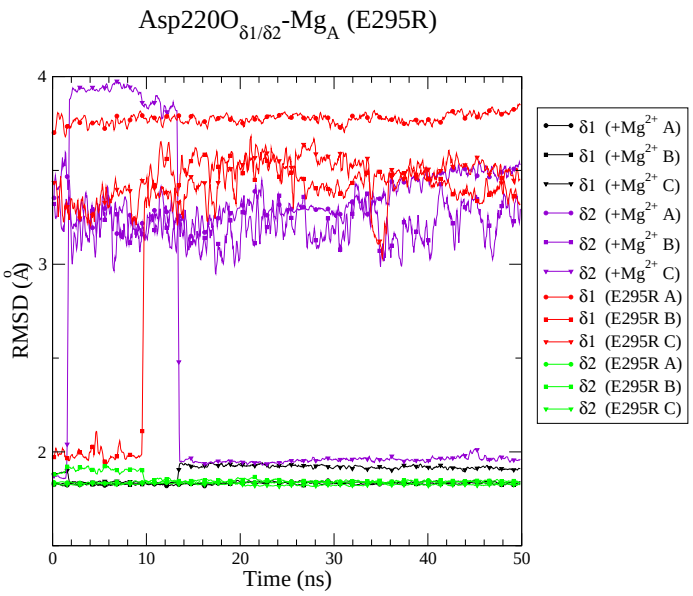


Figure A.17: Interaction between Asp 220 O and Mg_A^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

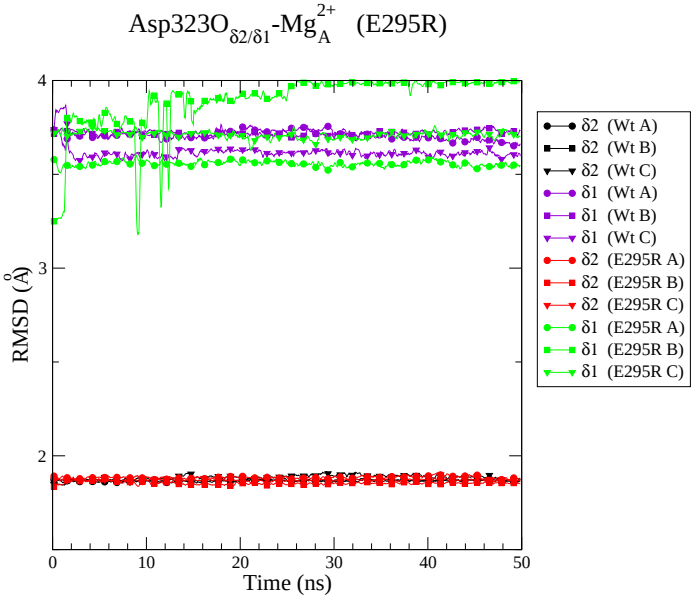


Figure A.18: Interaction between Asp 323 O and Mg_A^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

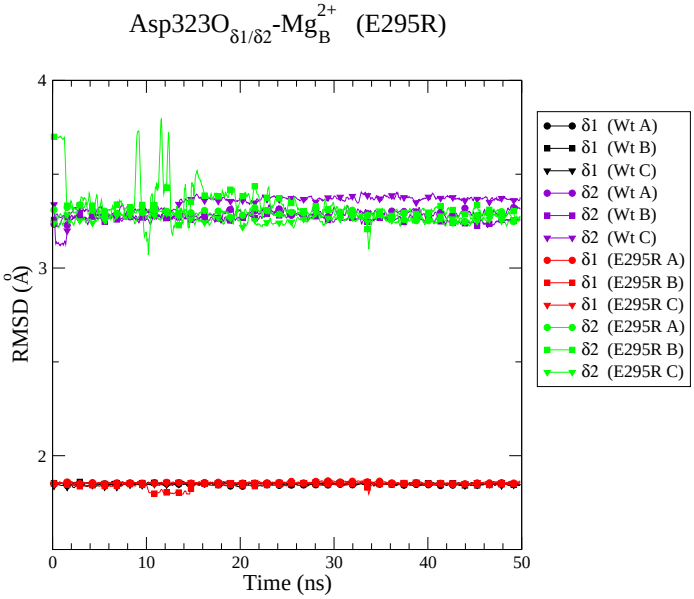


Figure A.19: Interaction between Asp 323 O and Mg_B^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

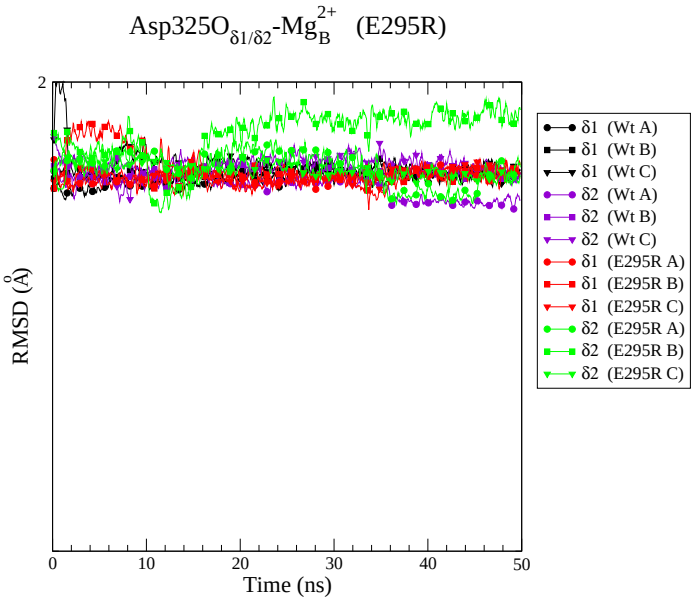


Figure A.20: Interaction between Asp 325 O and Mg_B^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

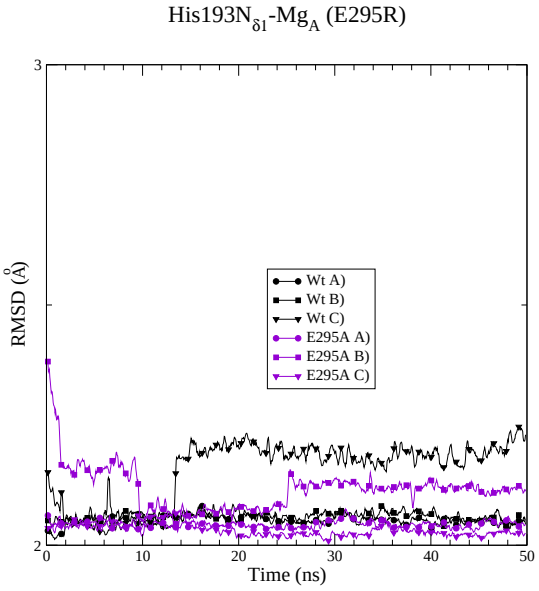


Figure A.21: Interaction between His 193 $N_{\delta1}$ and Mg_A^{2+} in *pf* Arg Glu 295 Arg compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

A.2.4 Glu 347 Gln

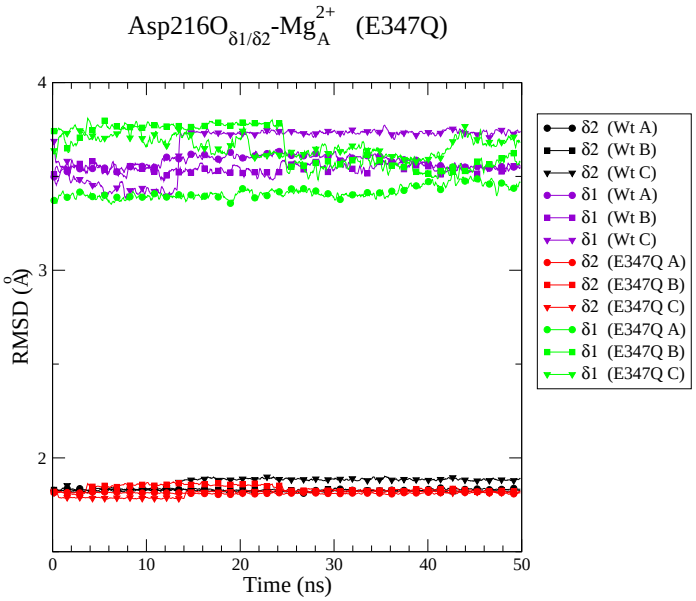


Figure A.22: Interaction between Asp 216 O and Mg^{2+}_{A} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

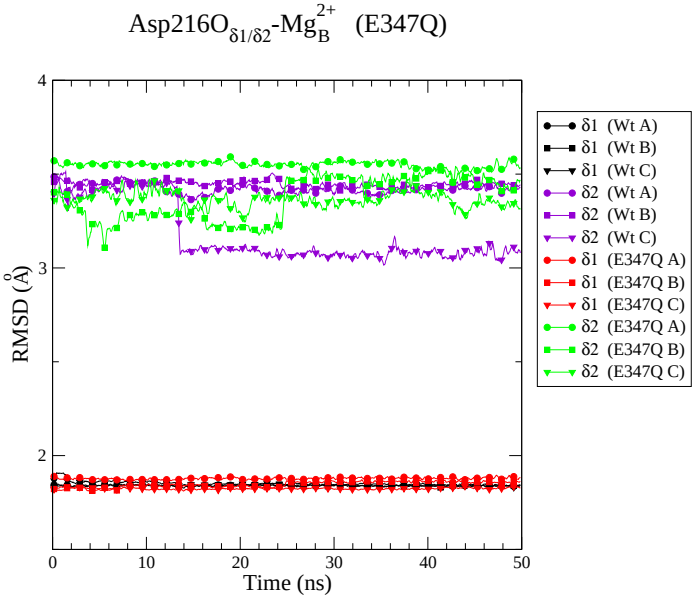


Figure A.23: Interaction between Asp 216 O and Mg^{2+}_{B} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

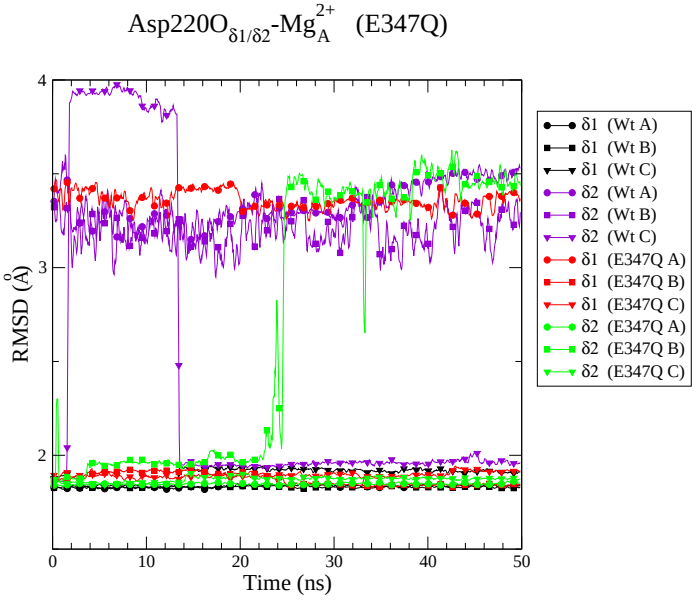


Figure A.24: Interaction between Asp 220 O and Mg^{2+}_{A} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

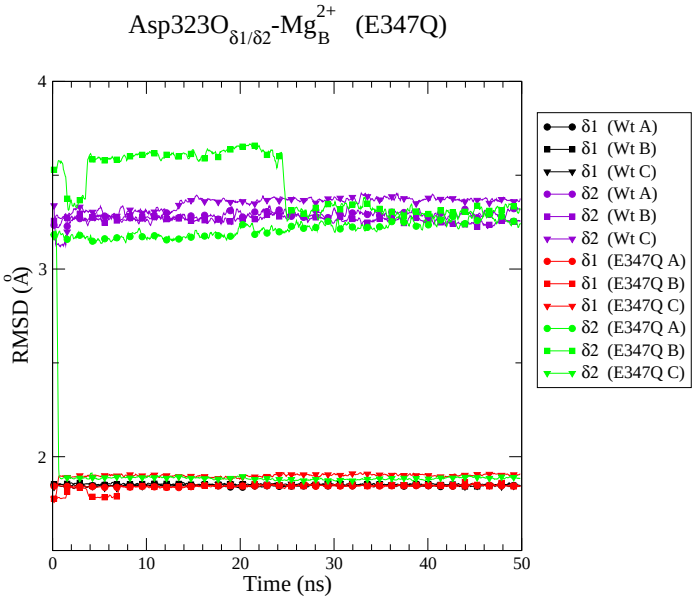


Figure A.25: Interaction between Asp 323 O and Mg_B^{2+} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

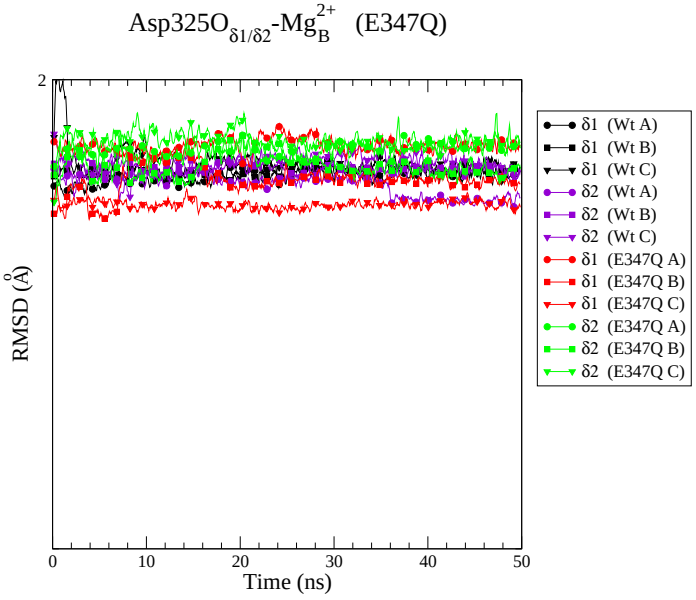


Figure A.26: Interaction between Asp 325 O and Mg_B^{2+} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

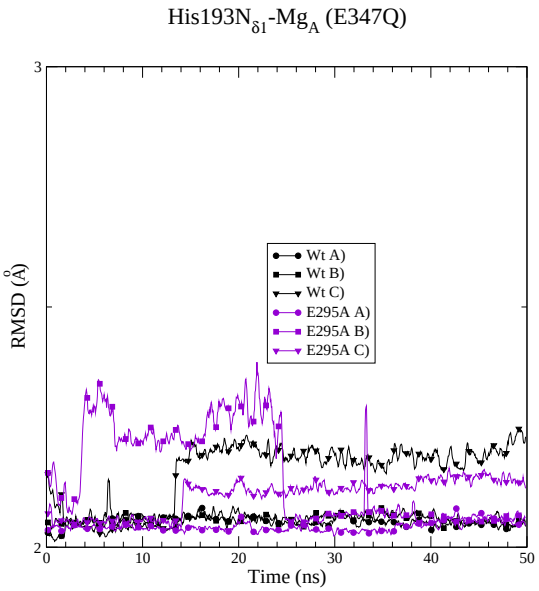


Figure A.27: Interaction between His 193 $N_{\delta1}$ and Mg_A^{2+} in *pf* Arg Glu 347 Gln compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

A.2.5 Arg 404 Ala

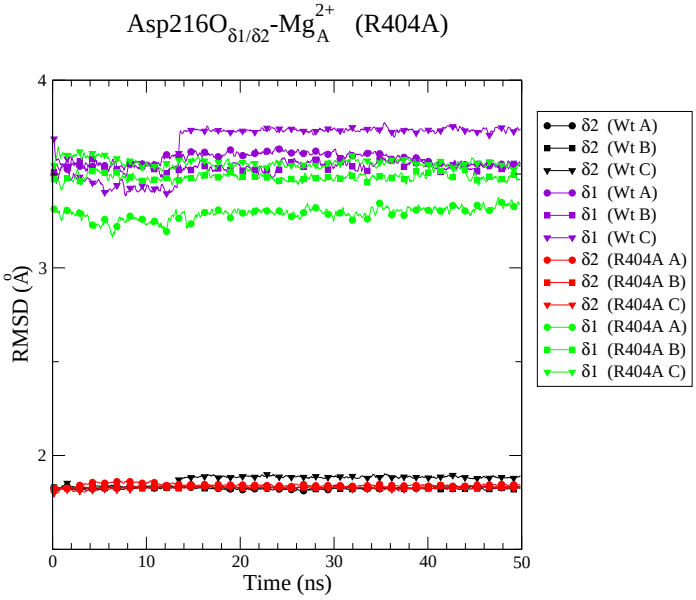


Figure A.28: Interaction between Asp 216 O and Mg^{2+}_{A} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

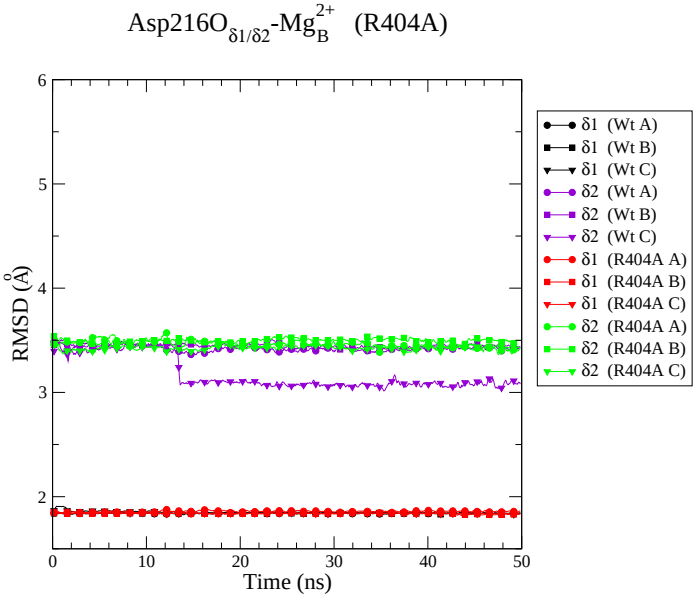


Figure A.29: Interaction between Asp 216 O and Mg^{2+}_{B} in *pf* Arg Arg,404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

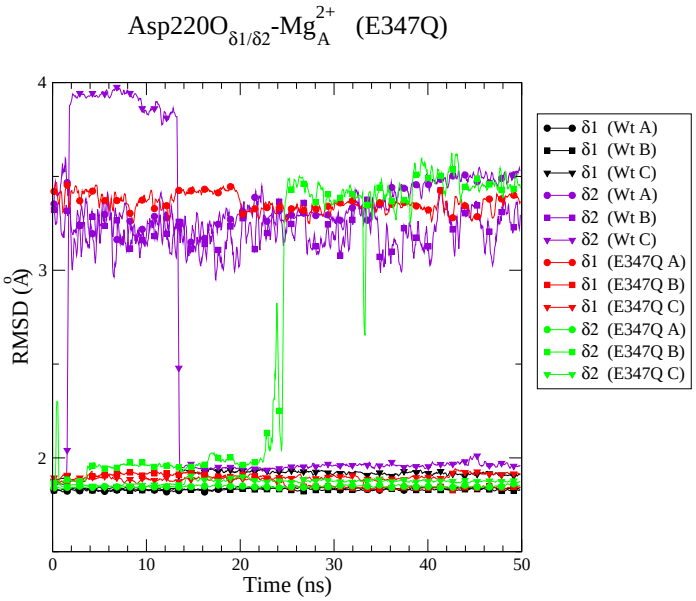


Figure A.30: Interaction between Asp 220 O and Mg^{2+}_{A} in *pf* Arg Arg,404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

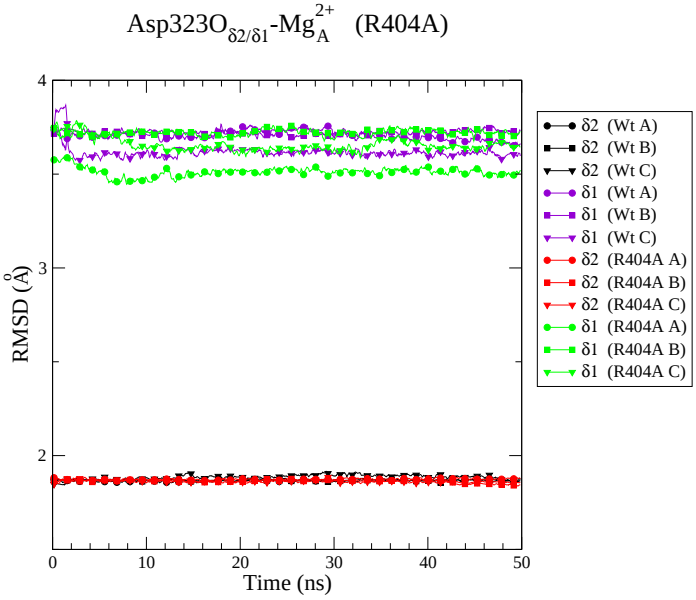


Figure A.31: Interaction between Asp 323 O and Mg^{2+}_{A} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

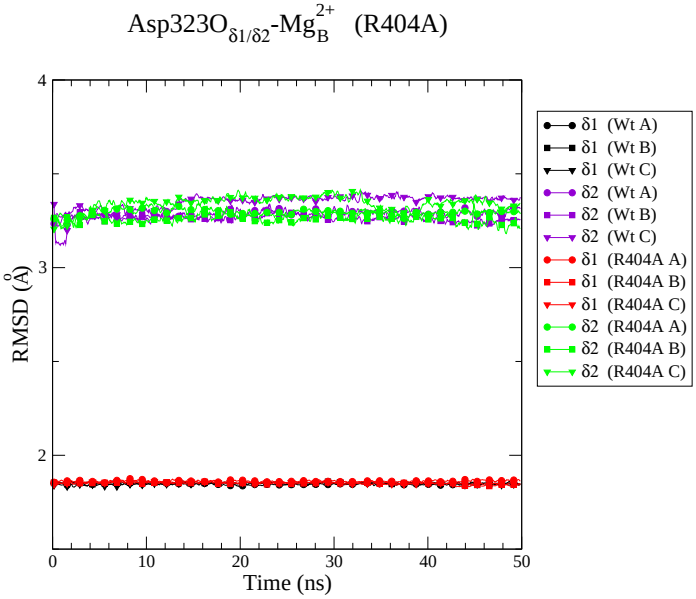


Figure A.32: Interaction between Asp 323 O and Mg^{2+}_{B} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

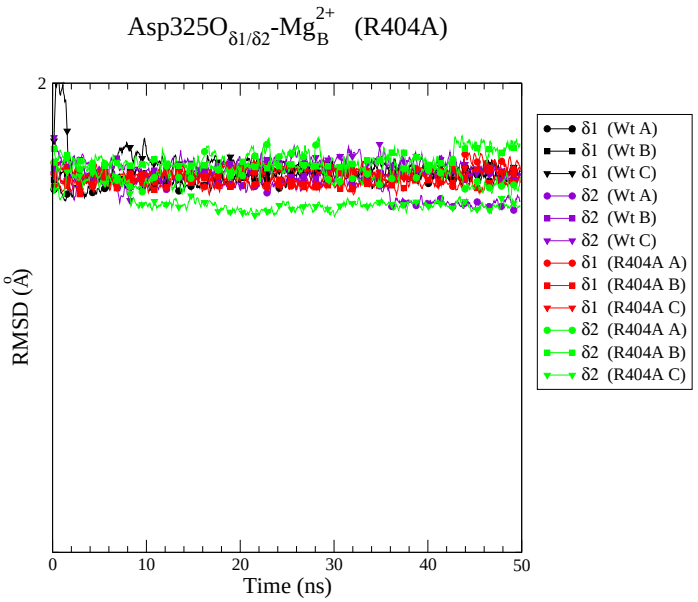


Figure A.33: Interaction between Asp 325 O and Mg_B^{2+} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

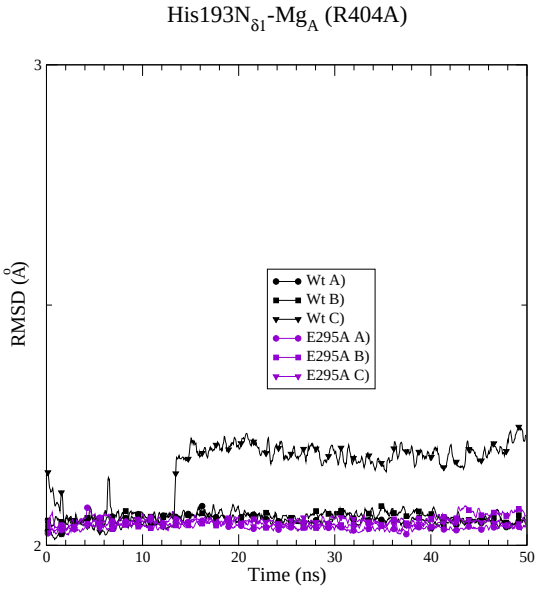


Figure A.34: Interaction between His 193 $N_{\delta1}$ and Mg_A^{2+} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

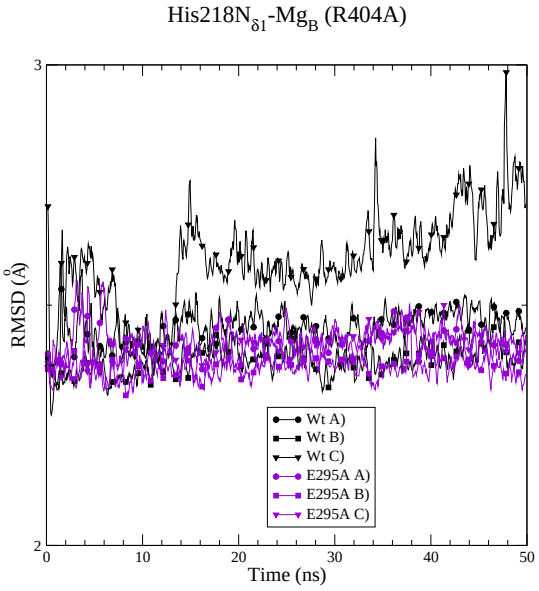


Figure A.35: Interaction between His 218 $N_{\delta1}$ and Mg_B^{2+} in *pf* Arg Arg 404 Ala compared to wild type. Both carboxyl oxygens ($\delta1/\delta2$) are included. Pairs of carboxyl O are indicated for chains A ($\circ$), B ($\square$) and C ($\triangledown$).

Appendix B

Supplementary data for Chapter 3

B.1 Model quality

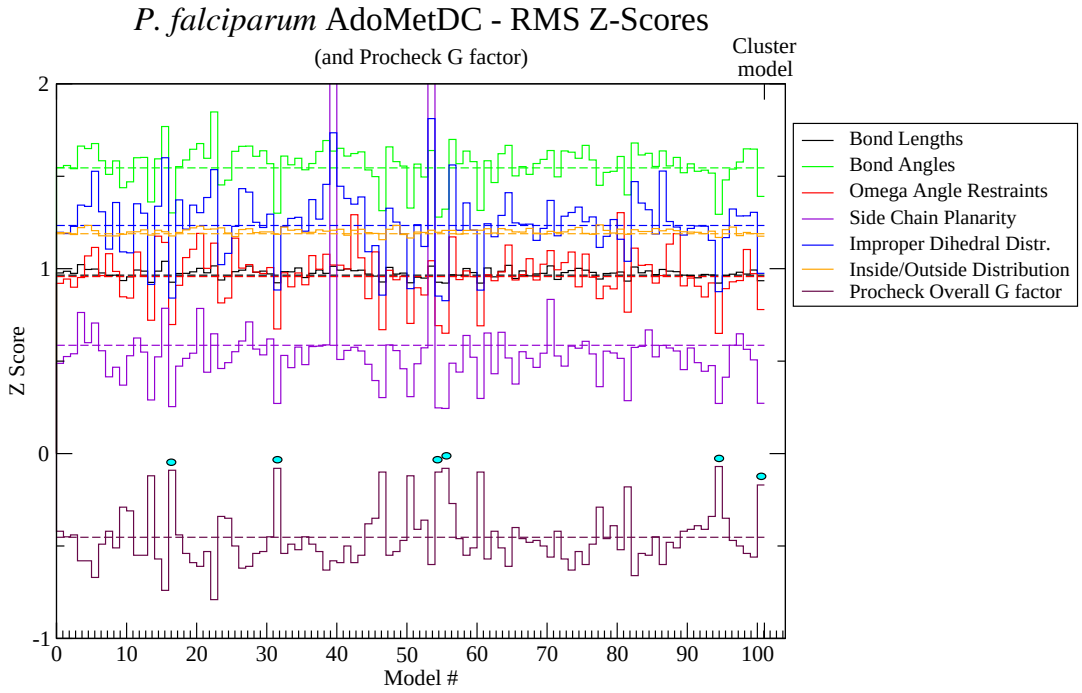


Figure B.1: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. falciparum* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

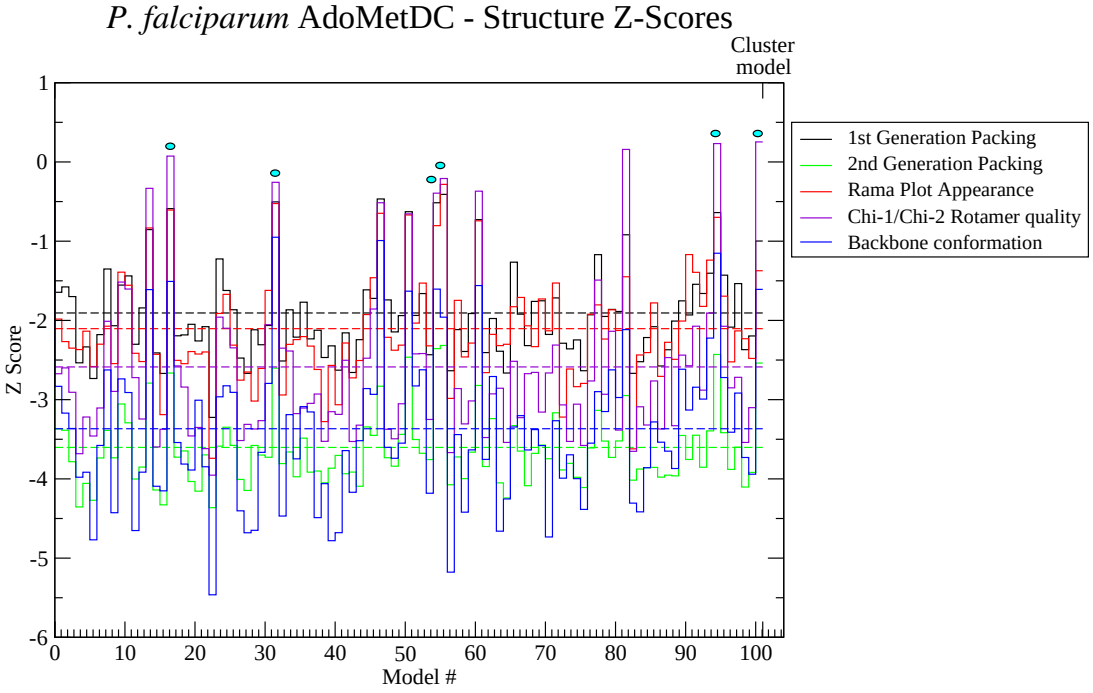


Figure B.2: WHATIF Structure Z-scores for modeling of *P. falciparum* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

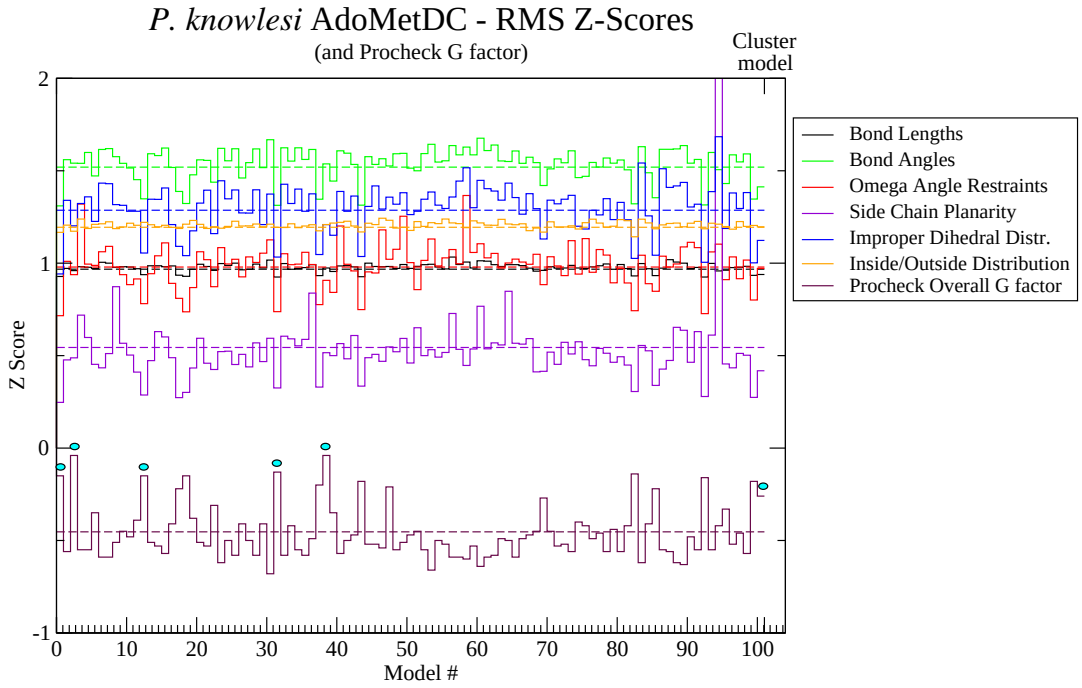


Figure B.3: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. knowlesi* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

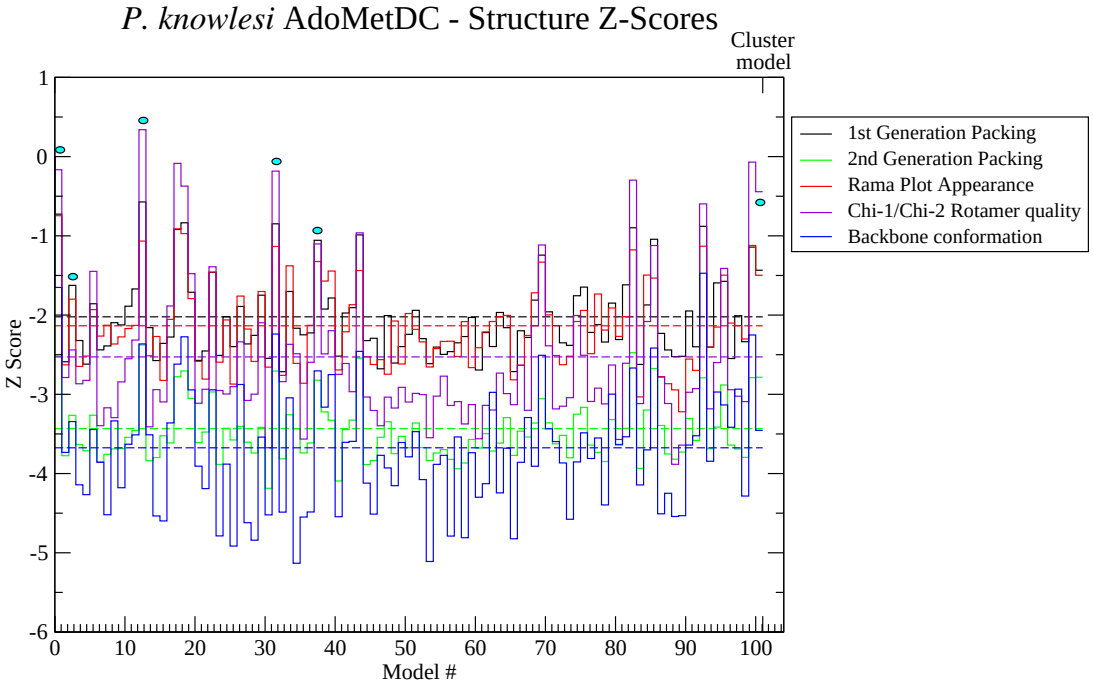


Figure B.4: WHATIF Structure Z-scores for modeling of *P. knowlesi* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

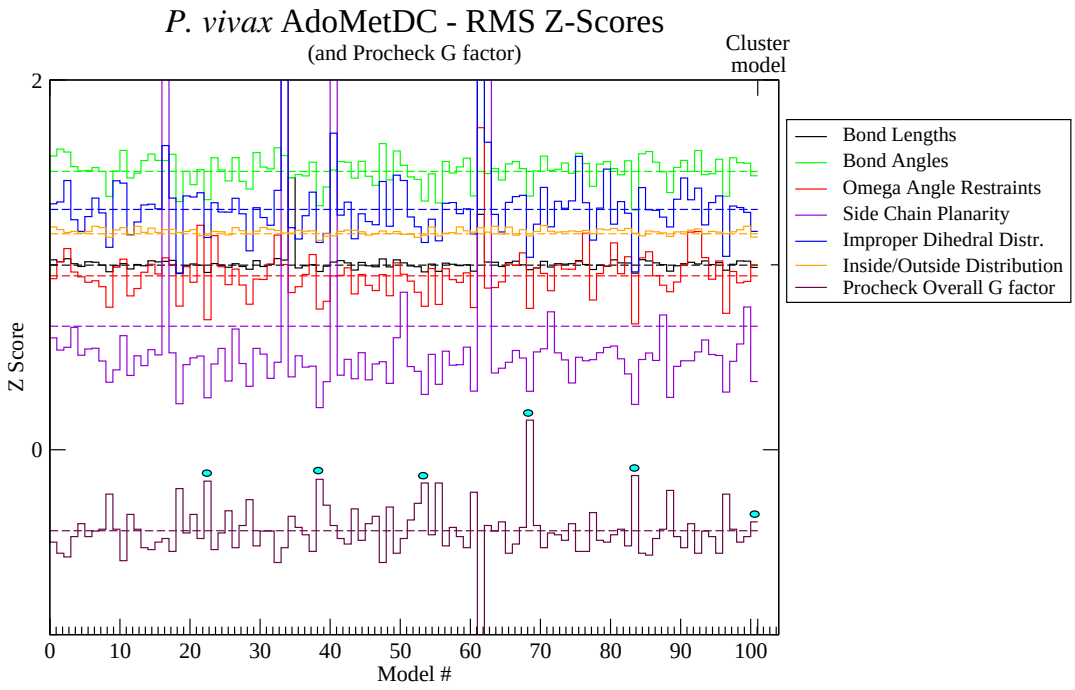


Figure B.5: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. vivax* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

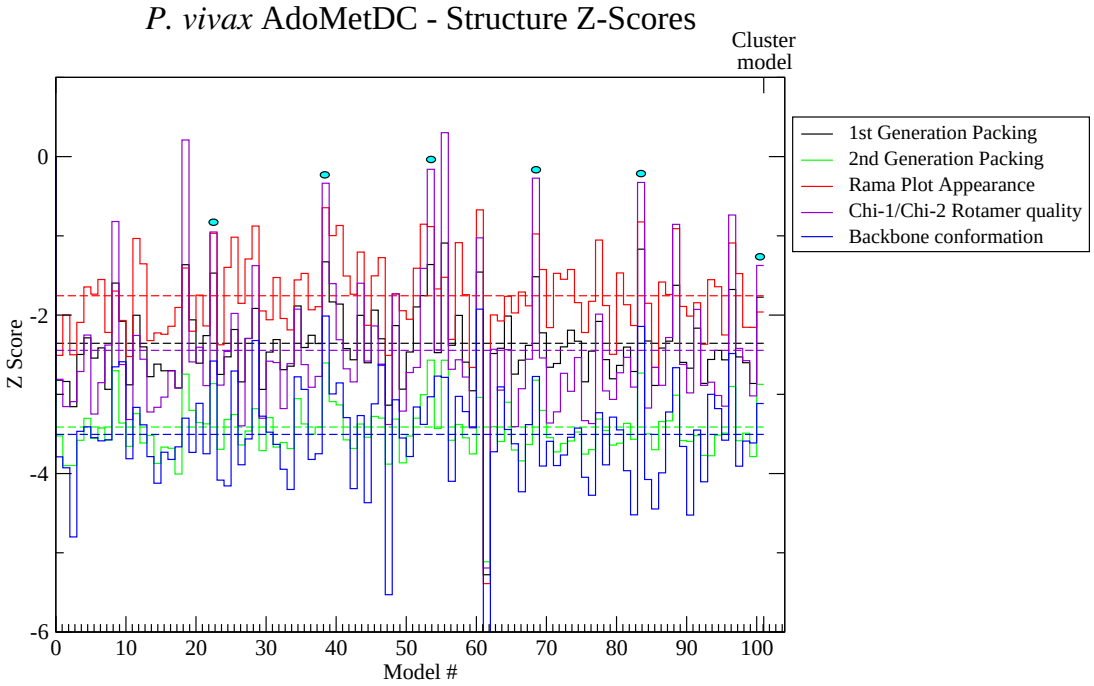


Figure B.6: WHATIF Structure Z-scores for modeling of *P. vivax* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

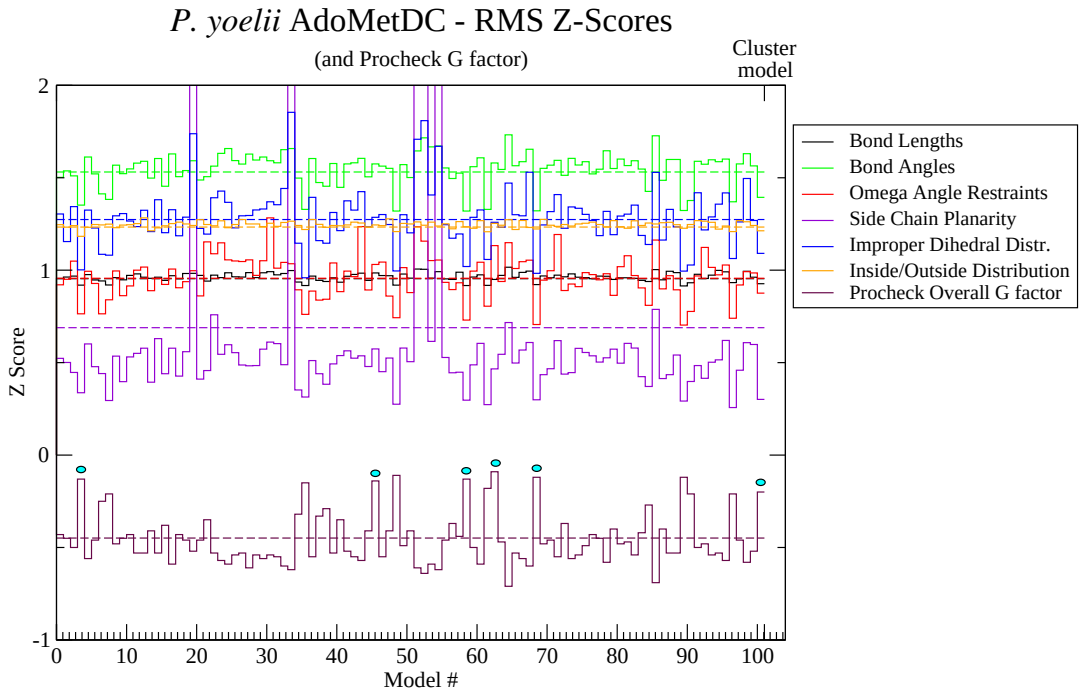


Figure B.7: WHATIF RMS Z-scores and PROCHECK G-factor for modeling of *P. yoelii* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

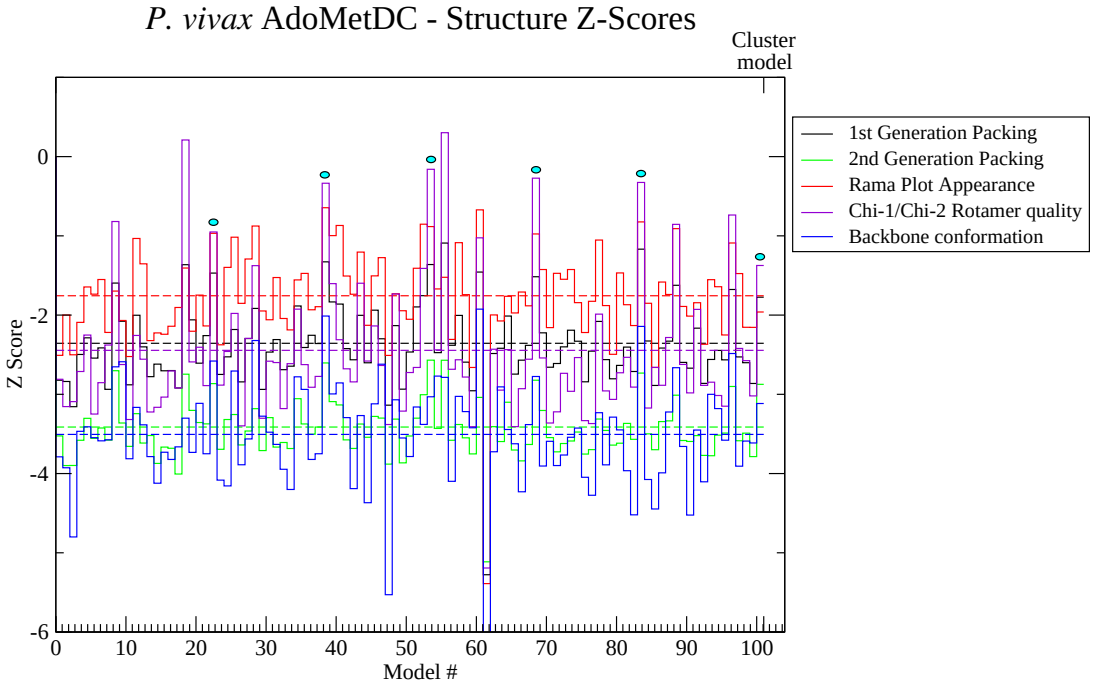


Figure B.8: WHATIF Structure Z-scores for modeling of *P. yoelii* AdoMetDC. The models chosen for docking are indicated by the cyan dots.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B.2 Surface distribution of divergence

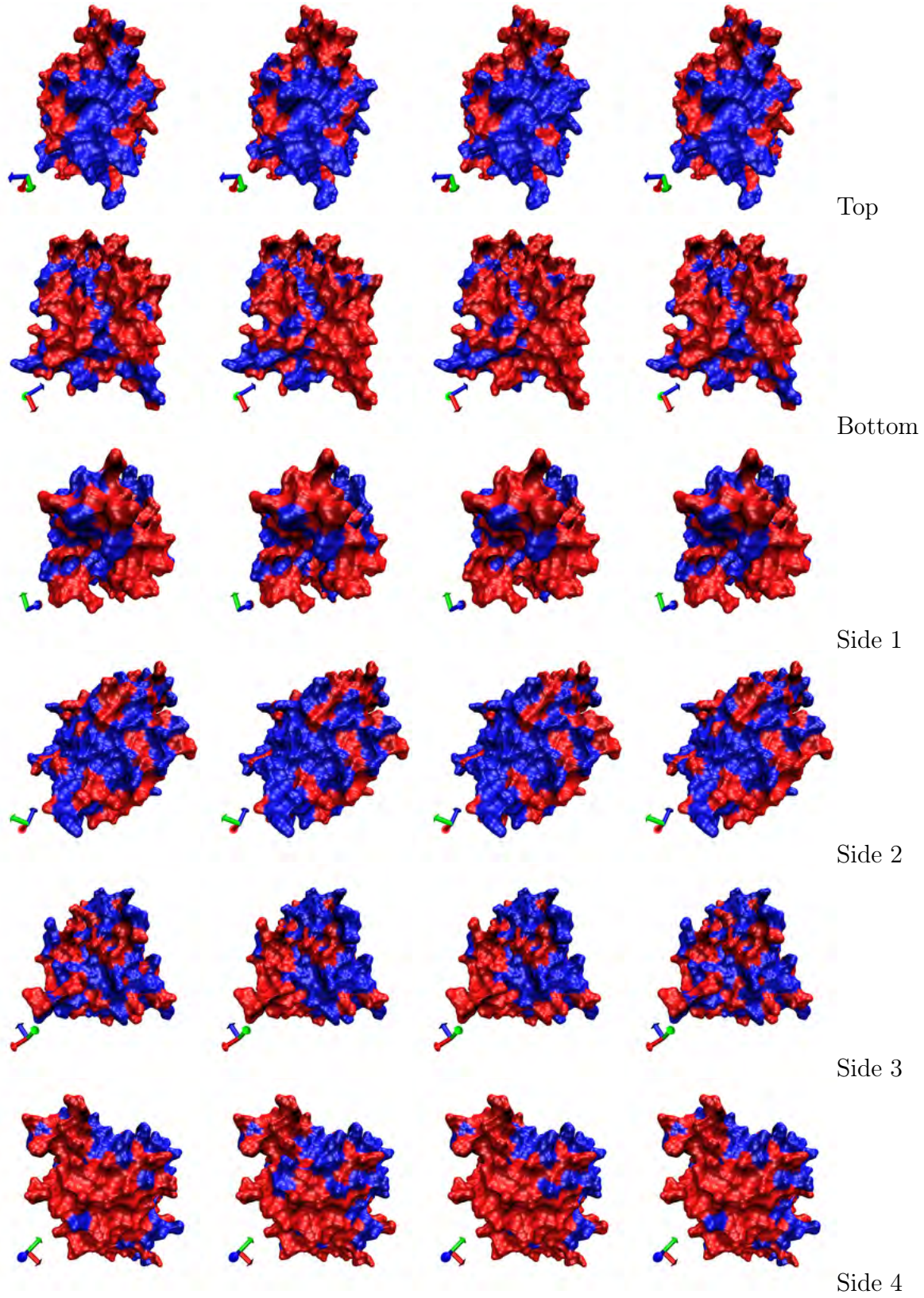


Figure B.9: Pairwise conservation of *P. falciparum* AdoMetDC surface residues. *P. falciparum* is compared to *P. berghei, P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

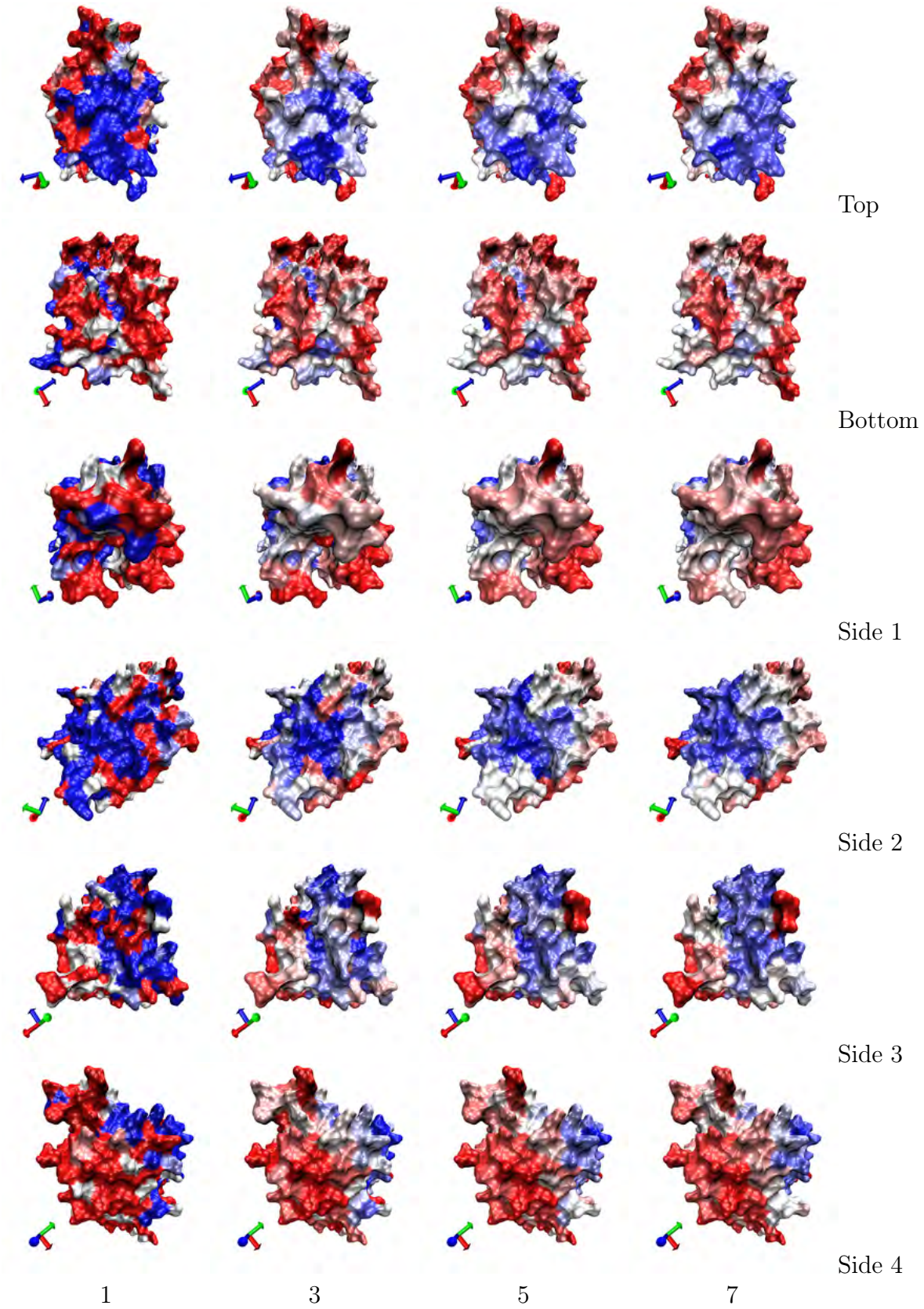


Figure B.10: Conservation of *P. falciparum* AdoMetDC surface residues. *P. falciparum* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

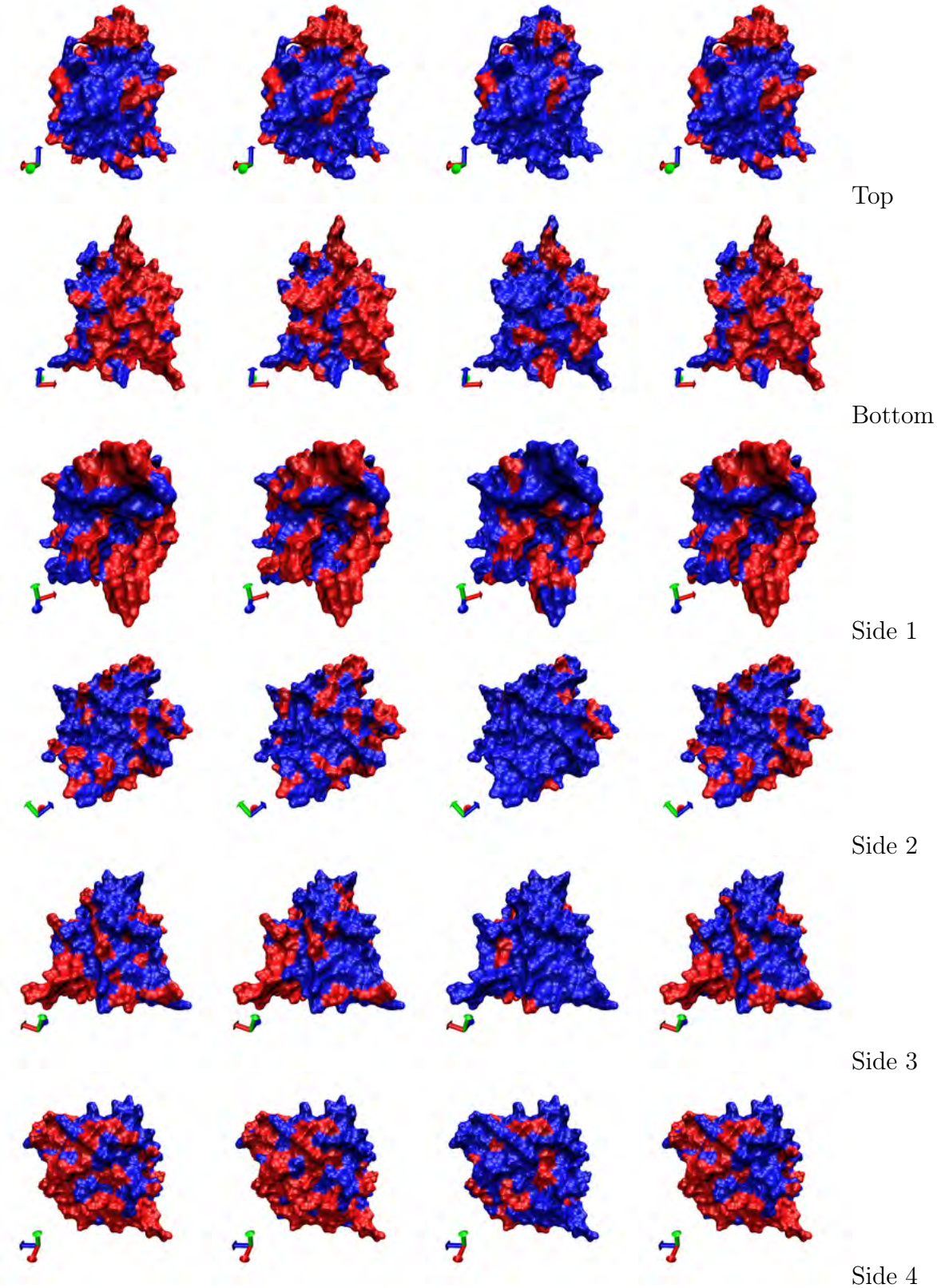


Figure B.11: Pairwise conservation of *P. knowlesi* AdoMetDC surface residues. *P. knowlesi* is compared to *P. berghei, P. falciparum*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

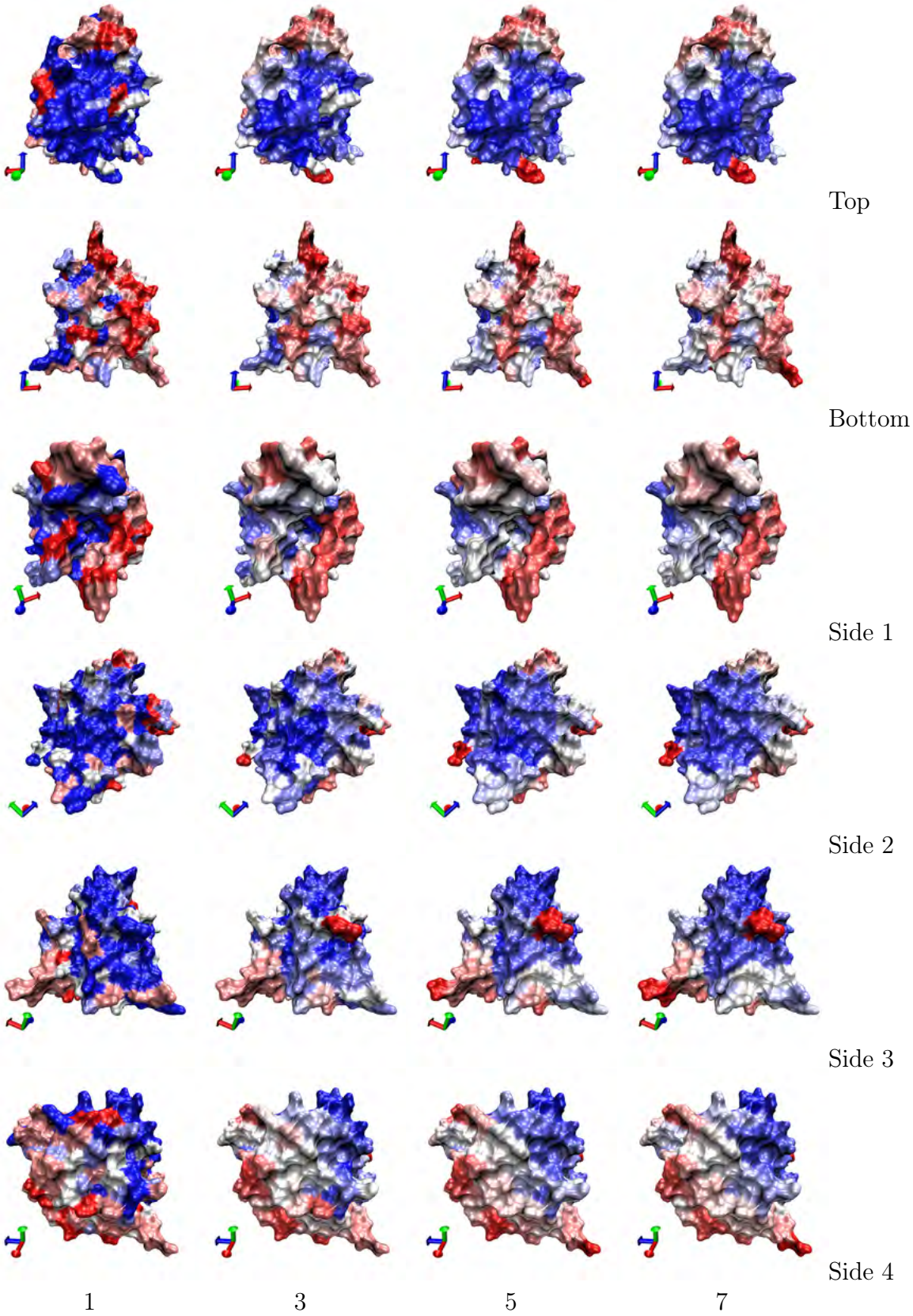


Figure B.12: Conservation of *P. knowlesi* AdoMetDC surface residues. *P. knowlesi* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

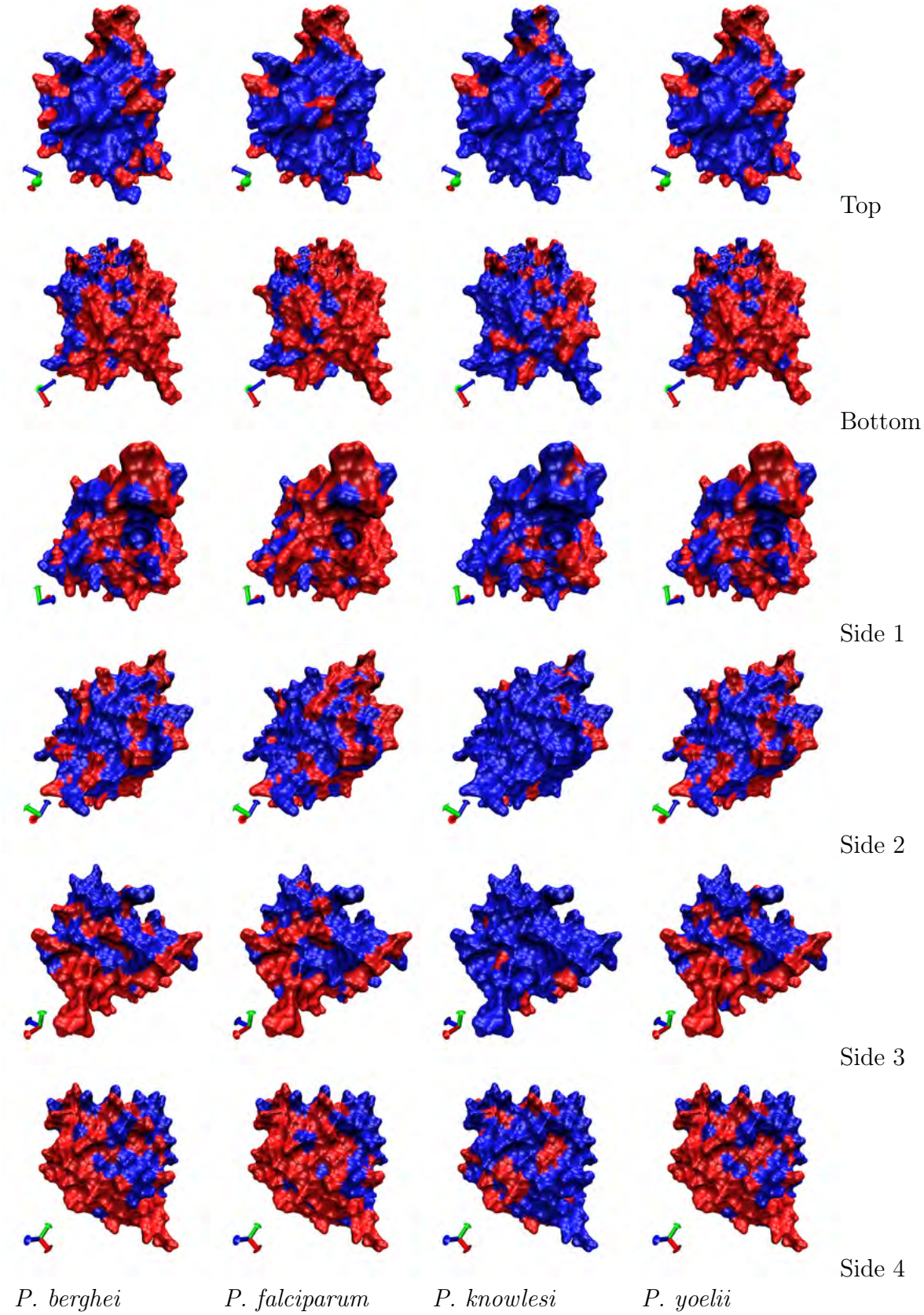


Figure B.13: Pairwise conservation of *P. vivax* AdoMetDC surface residues. *P. vivax* is compared to *P. berghei, P. falciparum*, *P. knowlesi* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

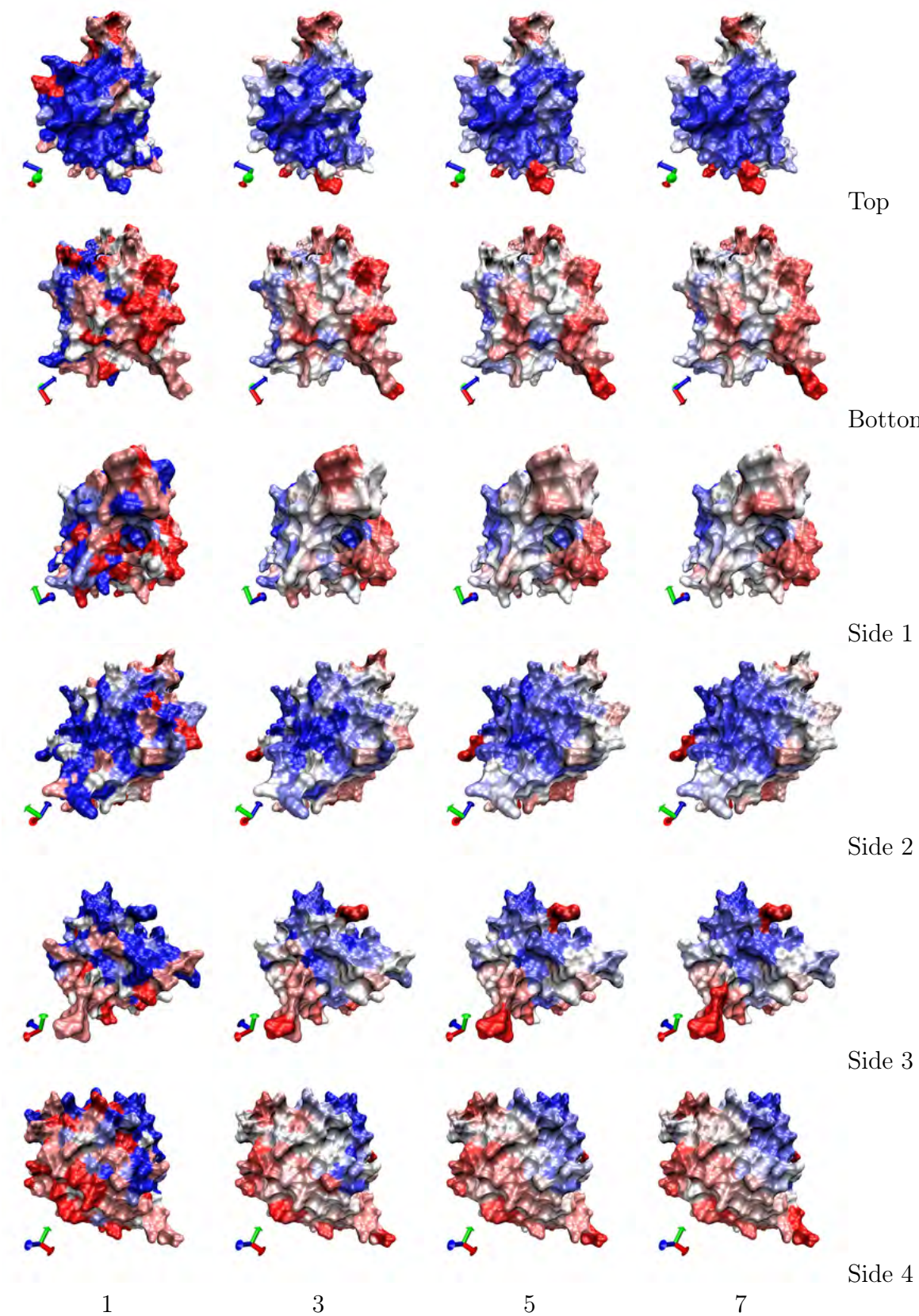


Figure B.14: Conservation of *P. vivax* AdoMetDC surface residues. *P. vivax* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

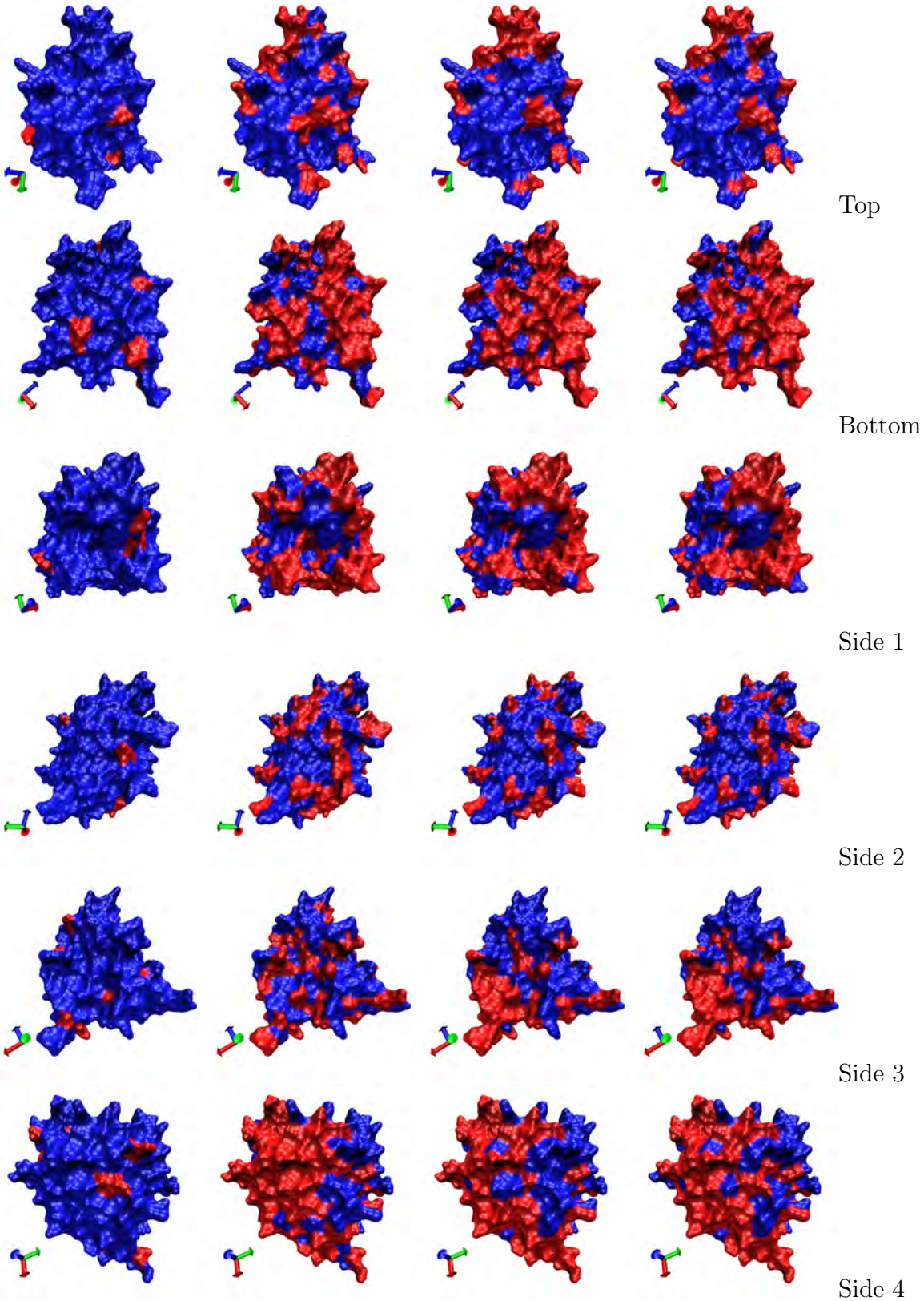


Figure B.15: Pairwise conservation of *P. yoelii* AdoMetDC surface residues. *P. yoelii* is compared to *P. berghei, P. falciparum*, *P. knowlesi* and *P. vivax* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

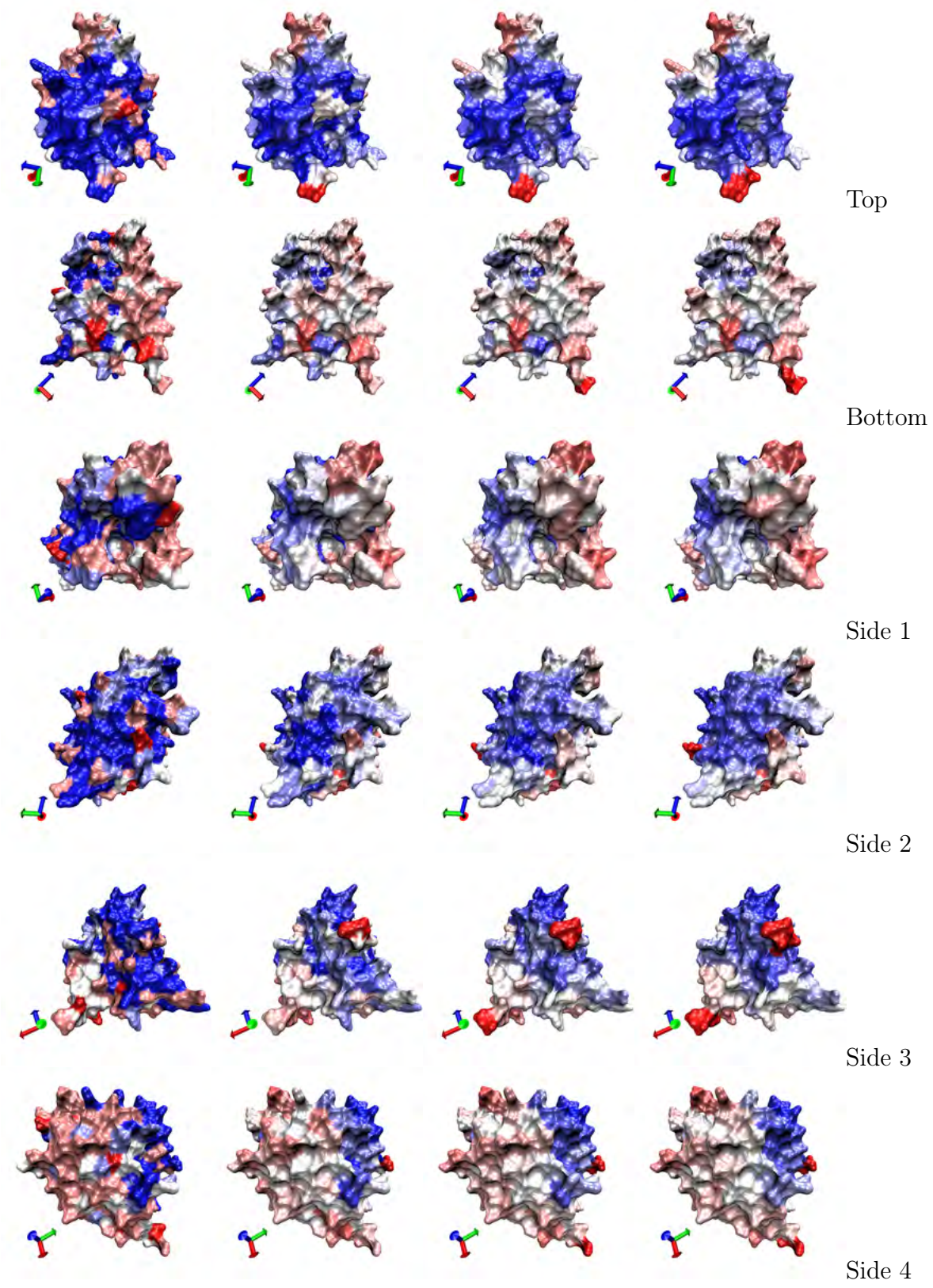


Figure B.16: Conservation of *P. yoelii* AdoMetDC surface residues. *P. yoelii* is compared simultaneously o all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

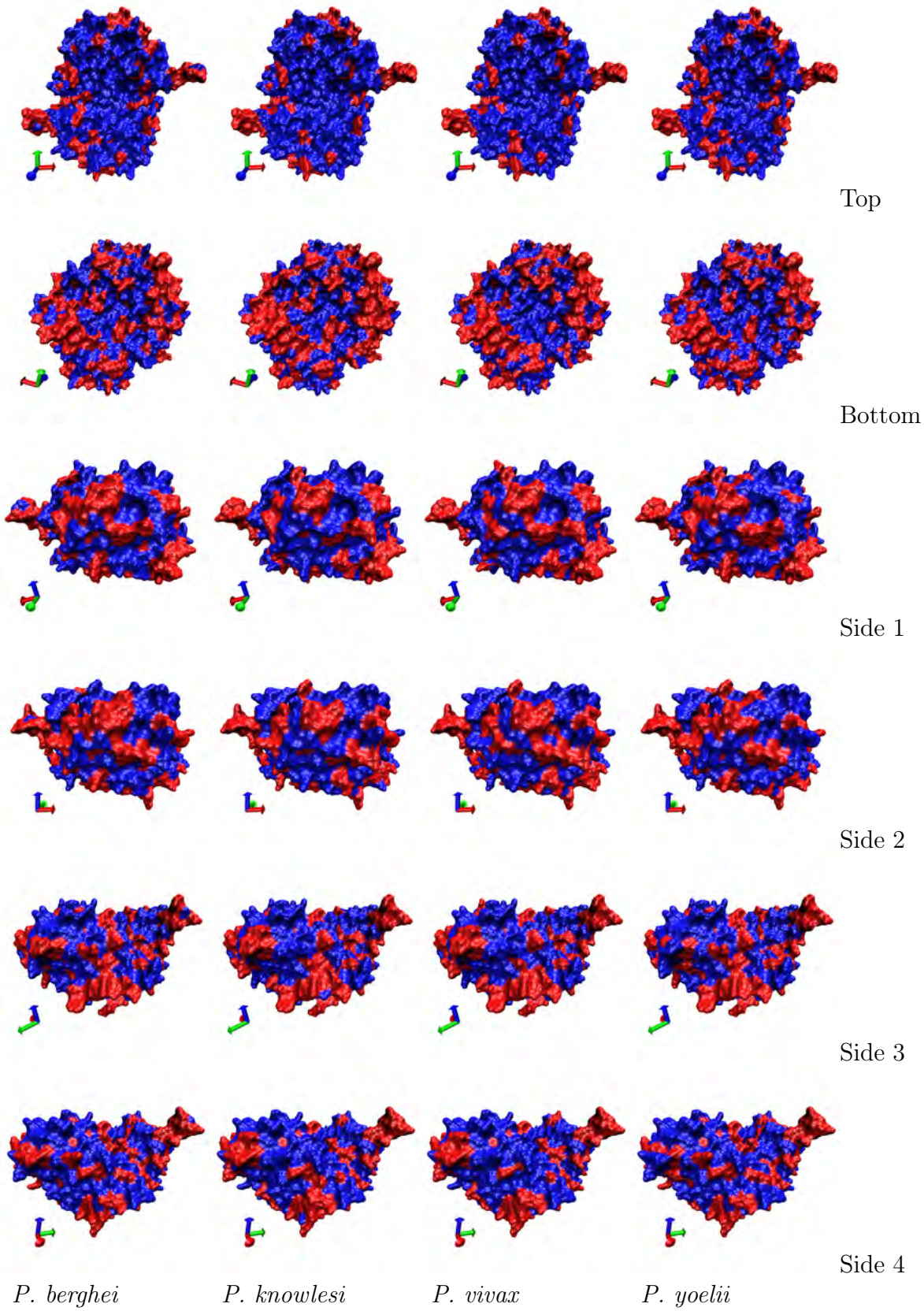


Figure B.17: Pairwise conservation of *P. falciparum* ODC surface residues. *P. falciparum* is compared to *P. berghei, P. knowlesi*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

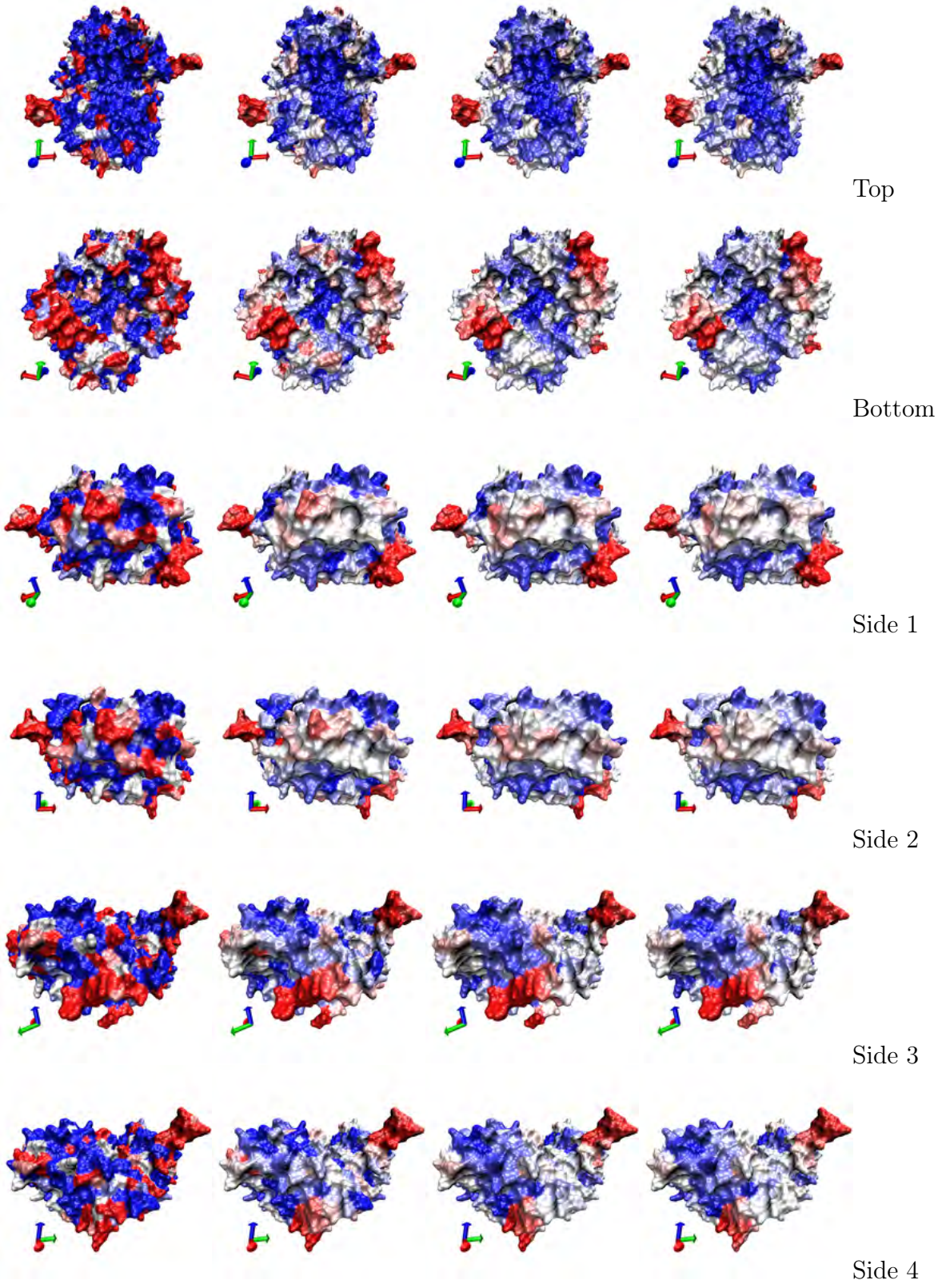


Figure B.18: Conservation of *P. falciparum* ODC surface residues. *P. falciparum* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

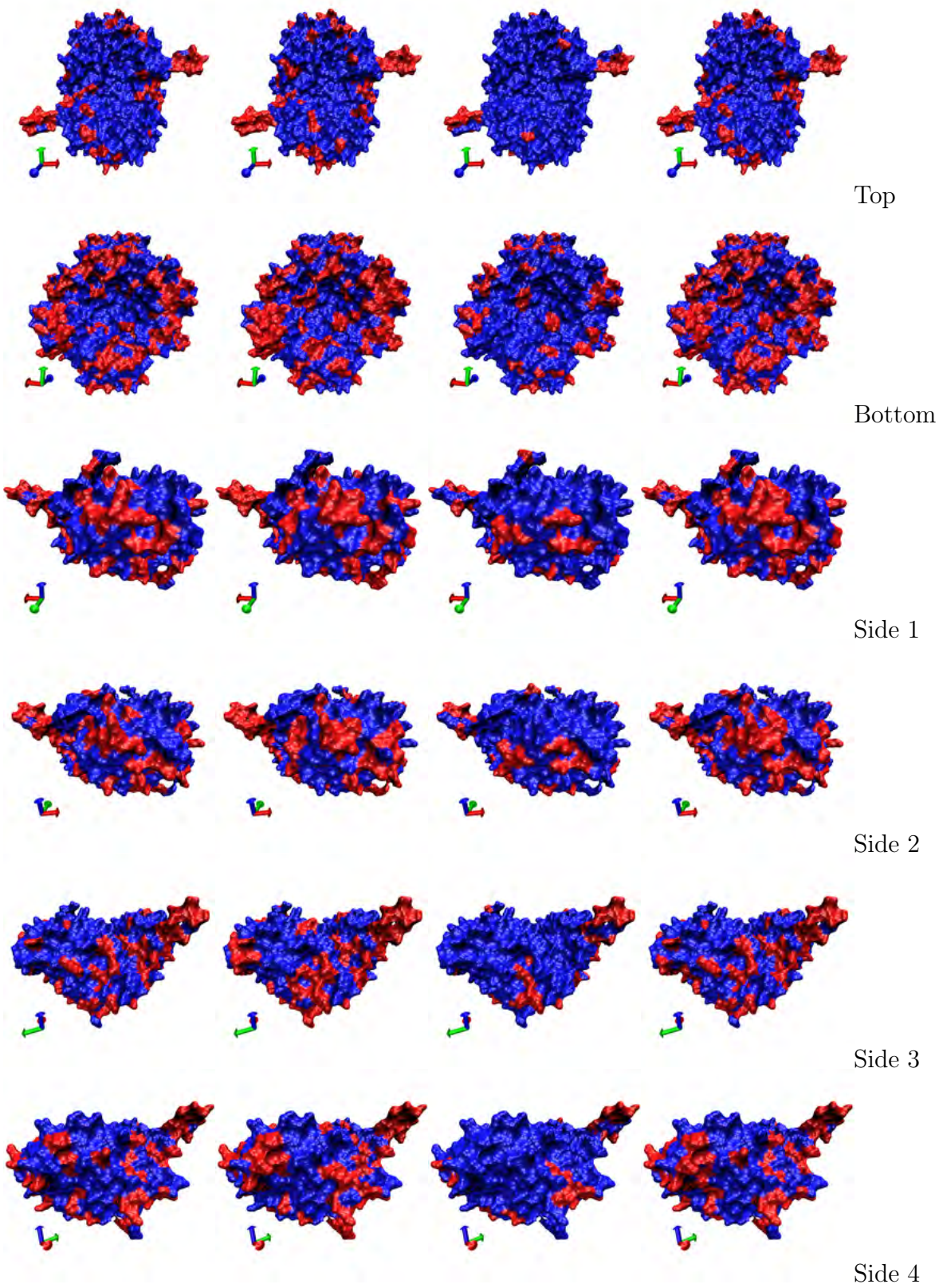


Figure B.19: Pairwise conservation of *P. knowlesi* ODC surface residues. *P. knowlesi* is compared to *P. berghei, P. falciparum*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

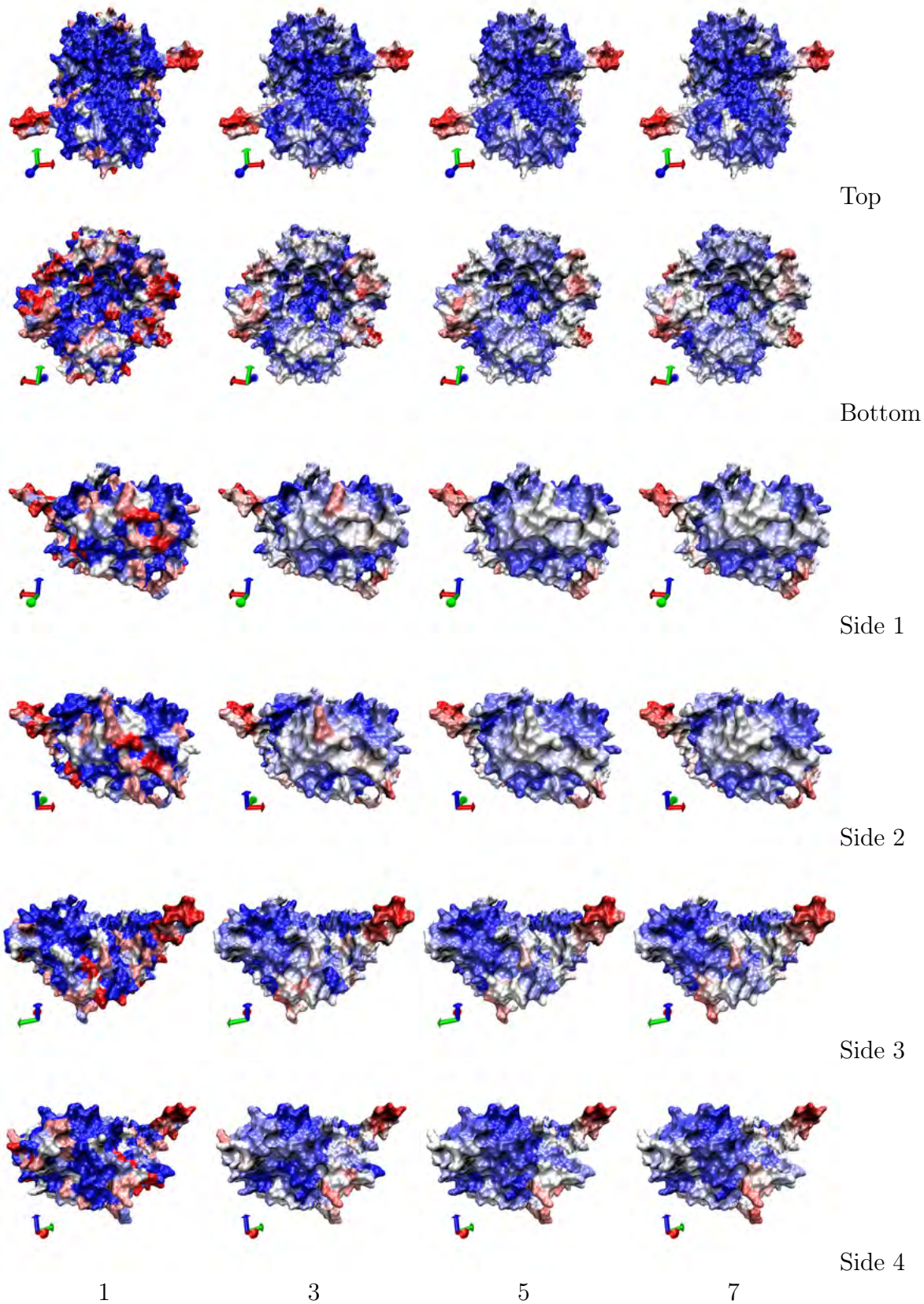


Figure B.20: Conservation of *P. knowlesi* ODC surface residues. *P. knowlesi* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

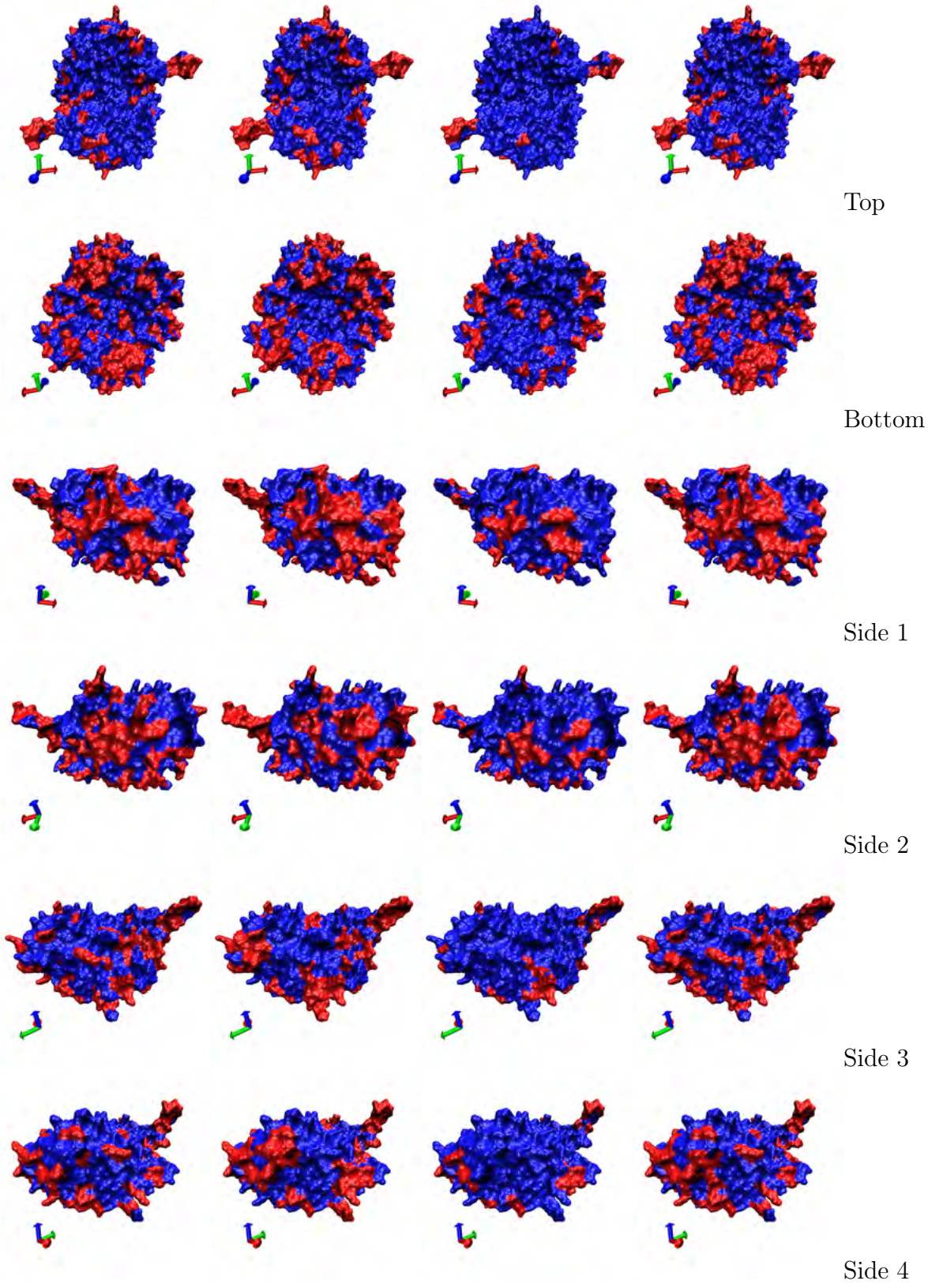


Figure B.21: Pairwise conservation of *P. vivax* ODC surface residues. *P. vivax* is compared to *P. berghei, P. falciparum*, *P. knowlesi* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

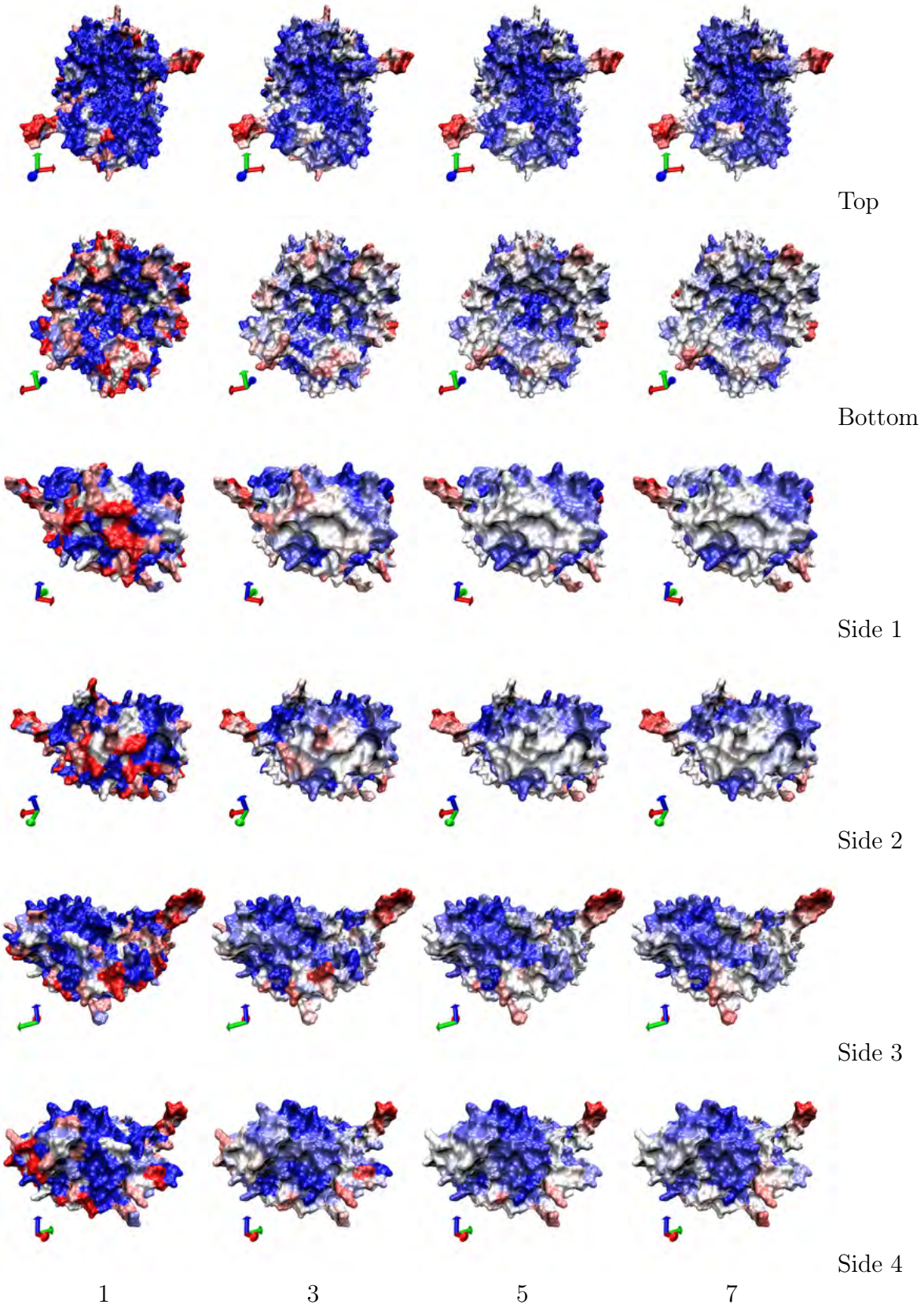


Figure B.22: Conservation of *P. vivax* ODC surface residues. *P. vivax* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

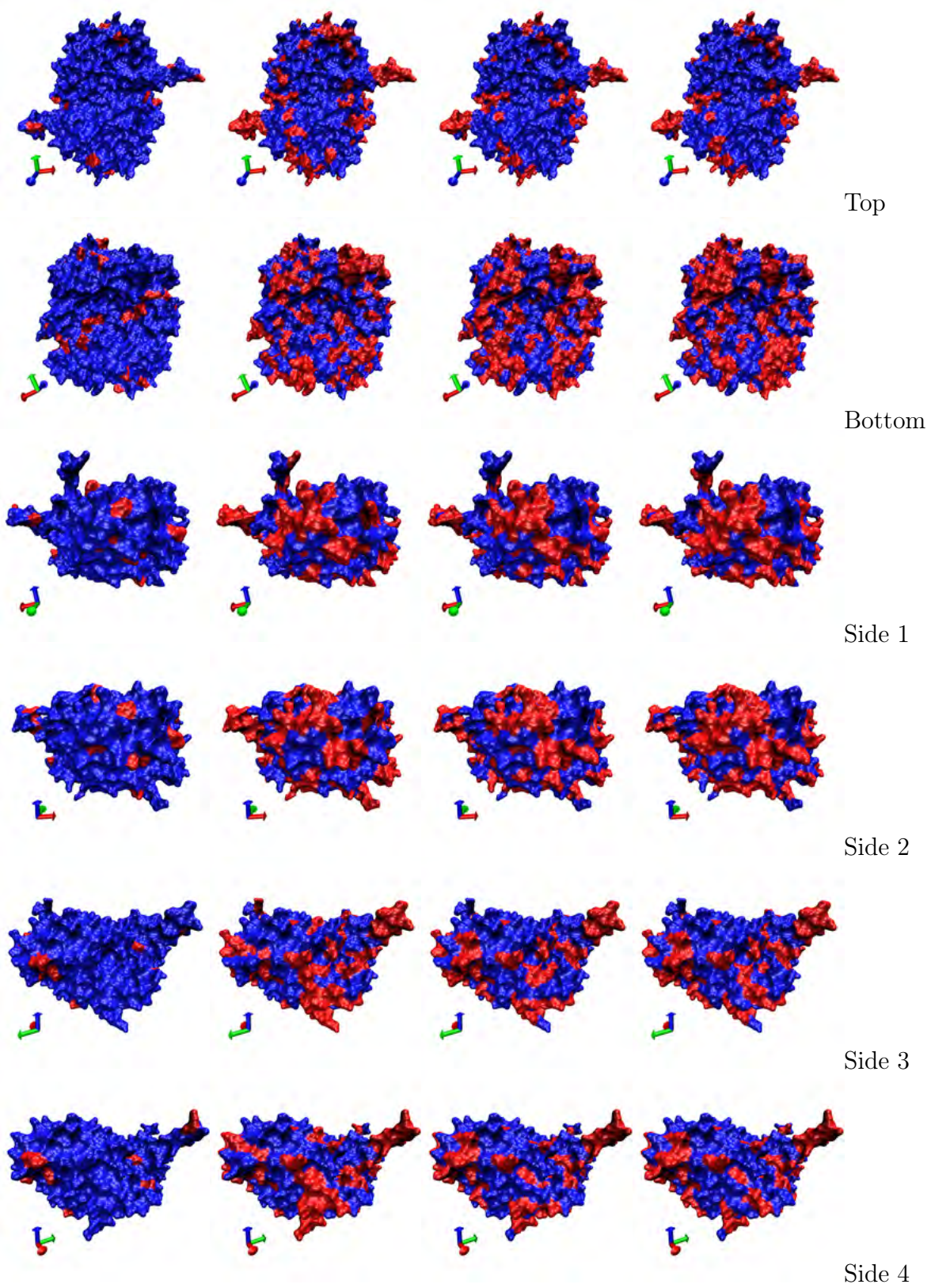


Figure B.23: Pairwise conservation of *P. yoelii* ODC surface residues. *P. yoelii* is compared to *P. berghei, P. falciparum*, *P. vivax* and *P. yoelii* in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

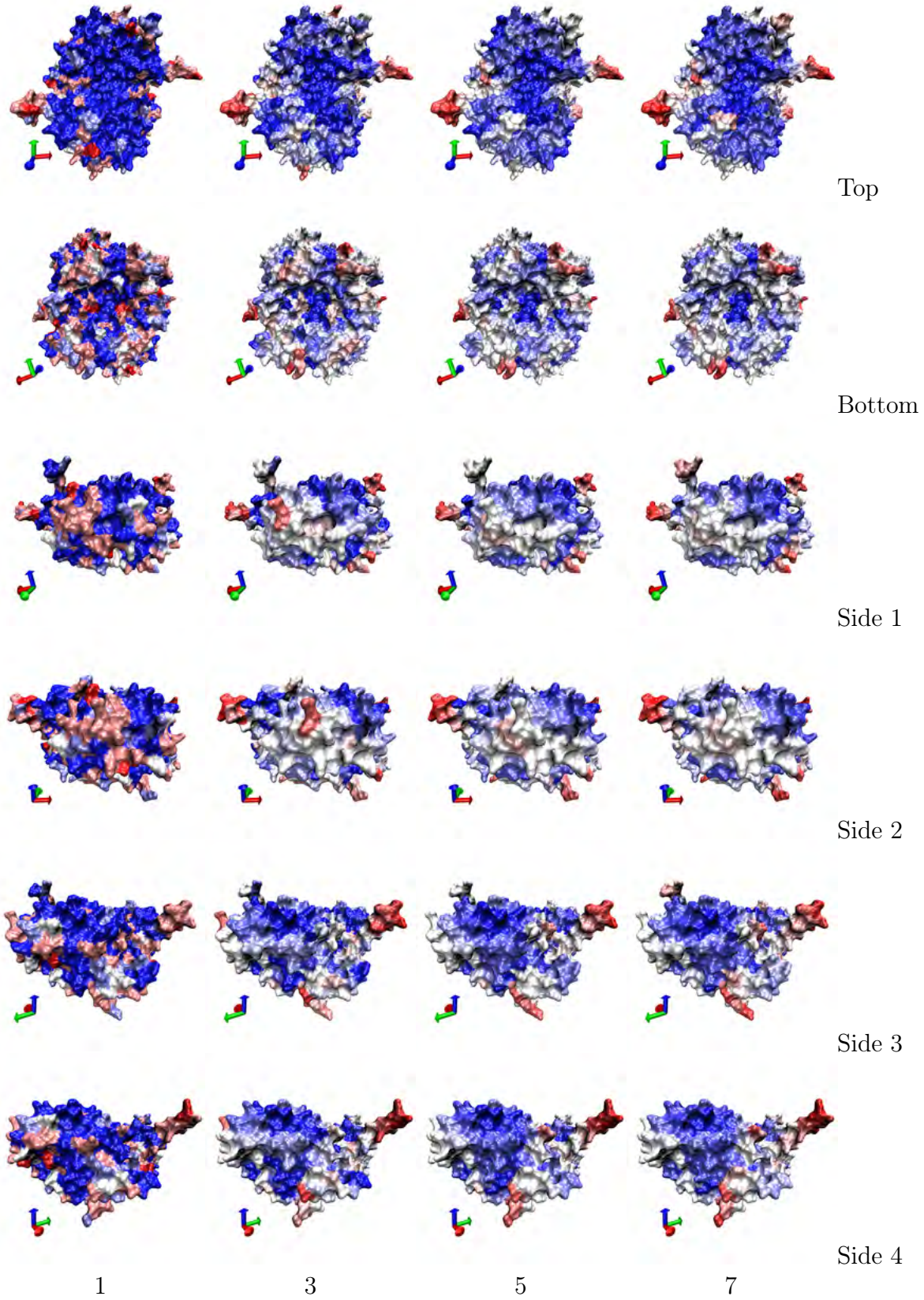


Figure B.24: Conservation of *P. yoelii* ODC surface residues. *P. yoelii* is compared simultaneously to all other *Plasmodium sp.* with sliding windows of 1, 3, 5 and 7 residues in columns 1-4, respectively. Rows 1 -6 correspond to the arbitrary top, bottom, side 1, side 2, side 3 and side 4 poses, respectively. Blue: identical residues, red: not conserved.

B.3 Distribution of RP scores

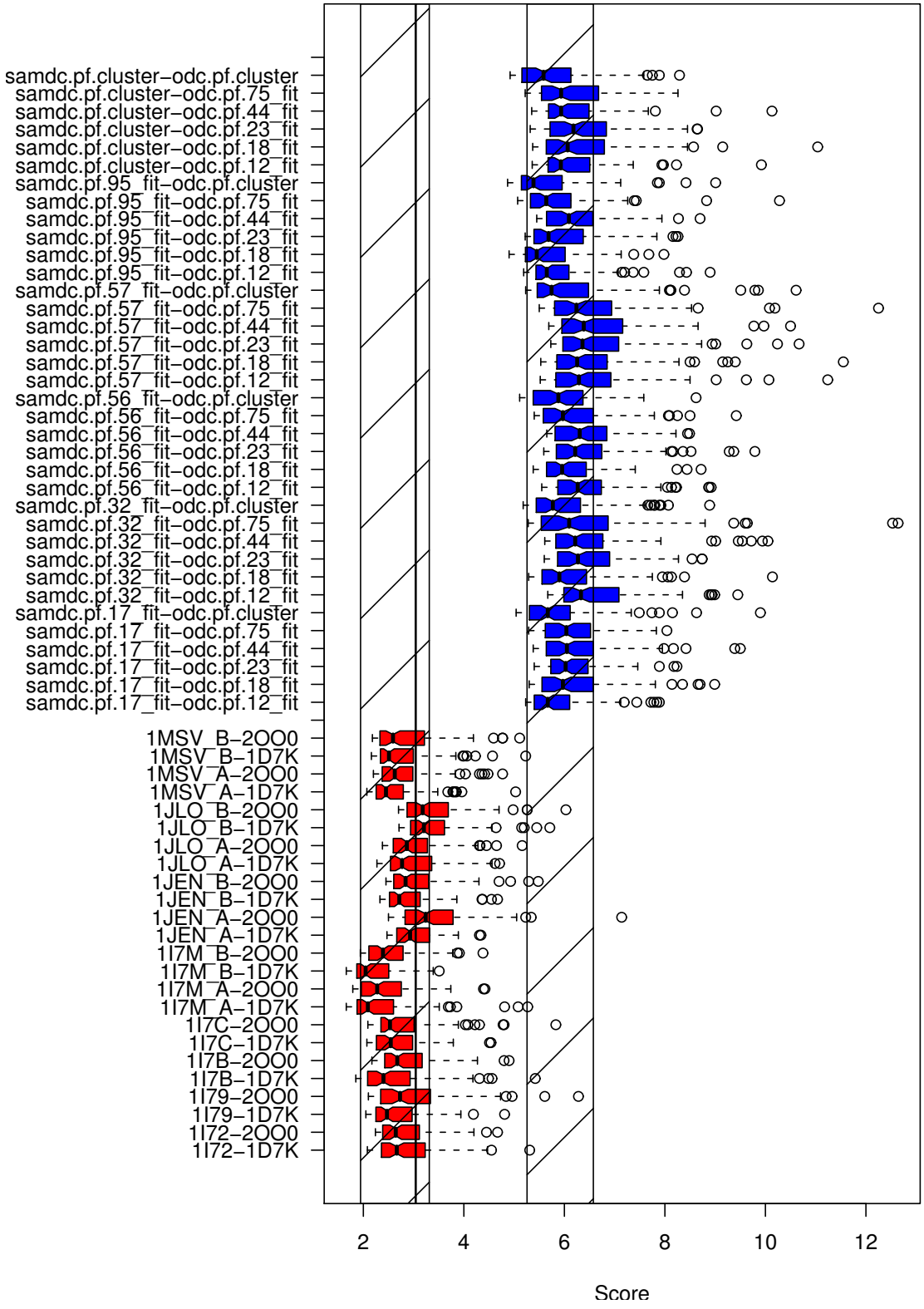


Figure B.25: Distribution of top 100 RP scores for docking of human (red) and *P. falciparum* (blue) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human and one for *P. falciparum* (The gap is too small to be visible on this scale). The RP score is plotted on the x-axis against the AdoMetDC/ODC model (blue) or structure (red) combinations.

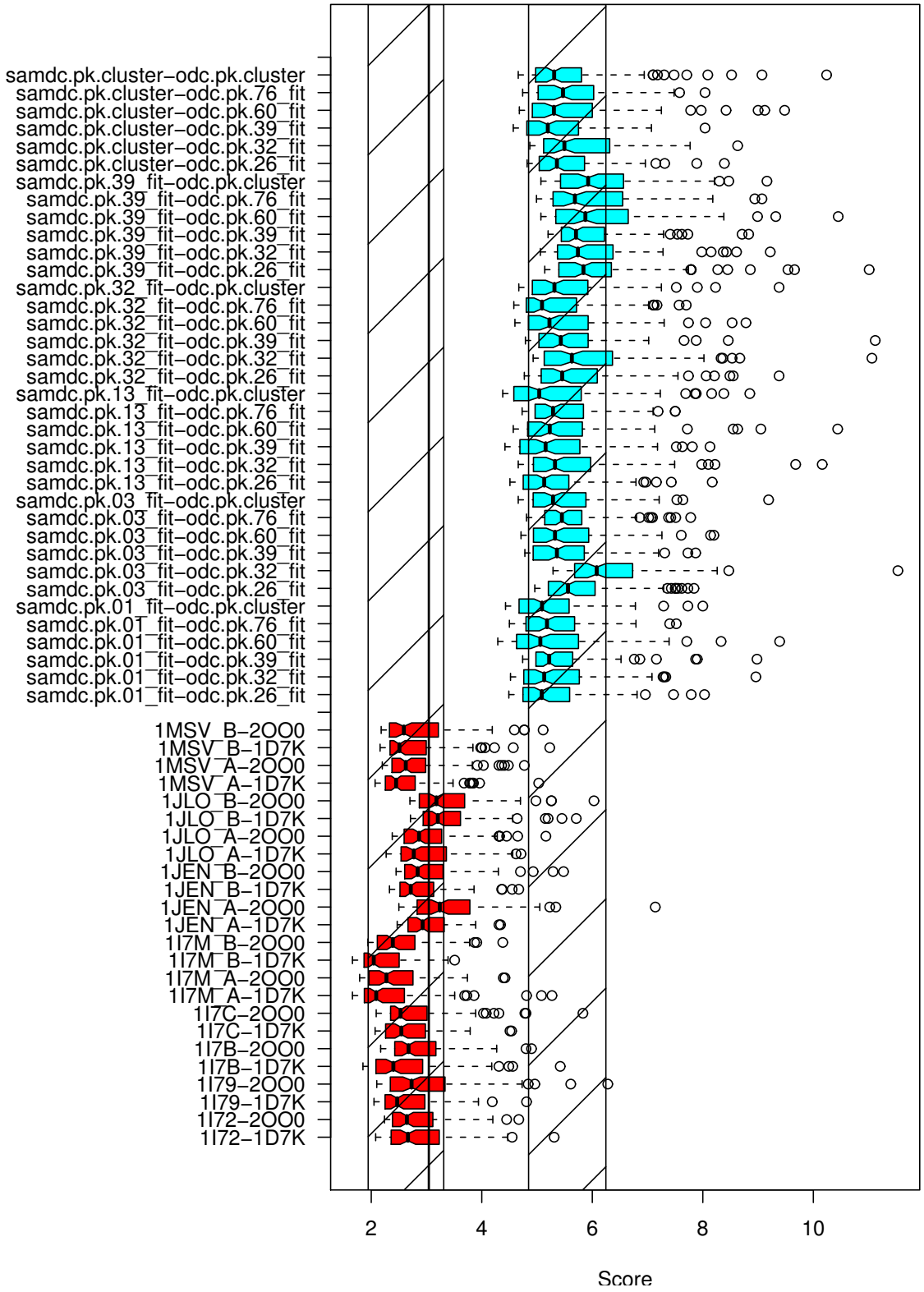


Figure B.26: Distribution of top 100 RP scores for docking of human (red) and *P. knowlesi* (cyan) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human and one for *P. knowlesi* (The gap is too small to be visible on this scale). The RP score is plotted on the x-axis against the AdoMetDC/ODC model (cyan) or structure (red) combinations.

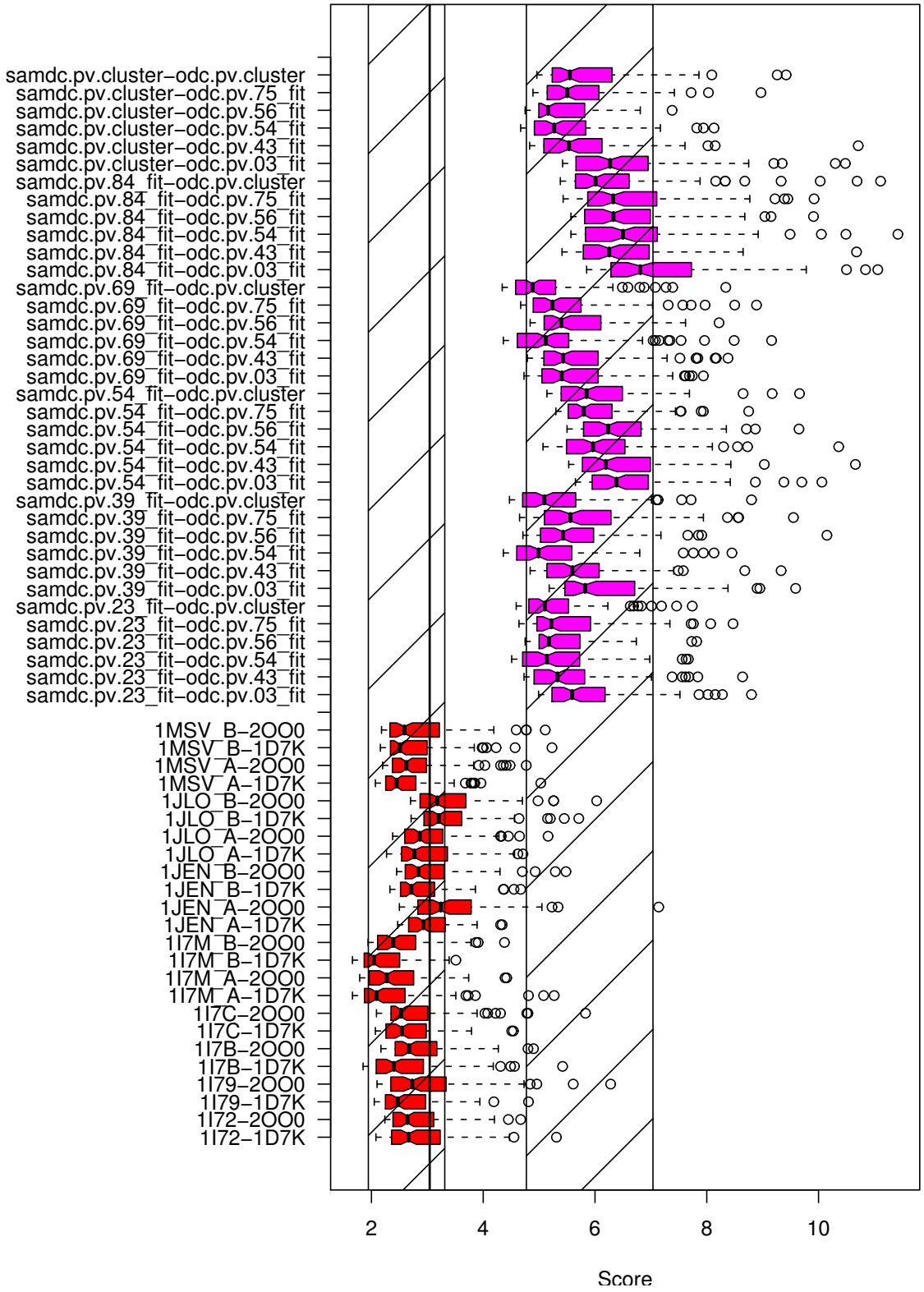


Figure B.27: Distribution of top 100 RP scores for docking of human (red) and *P. vivax* (magenta) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human and one for *P. vivax* (The gap is too small to be visible on this scale). The RP score is plotted on the x-axis against the AdoMetDC/ODC model (pink) or structure (red) combinations.

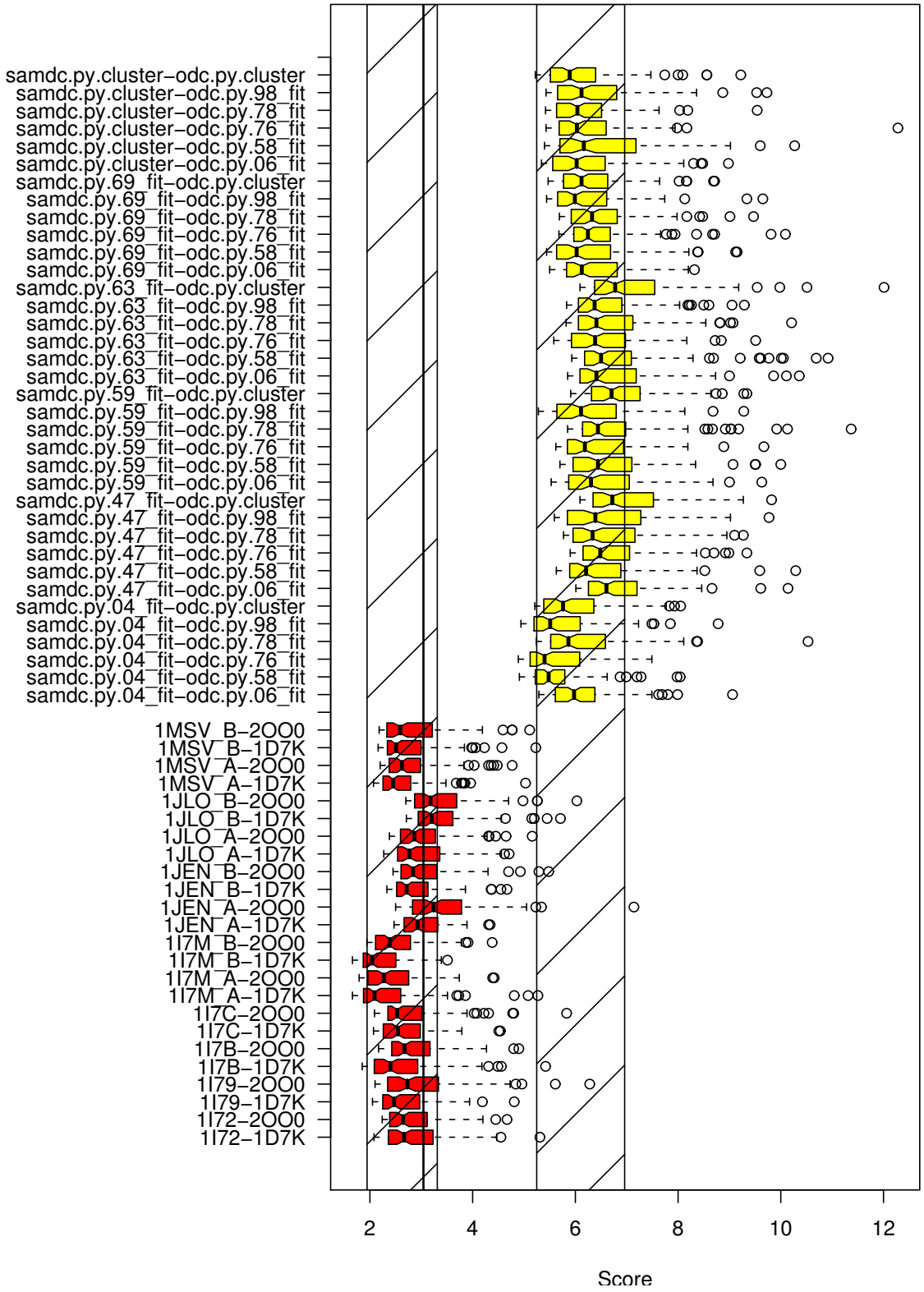
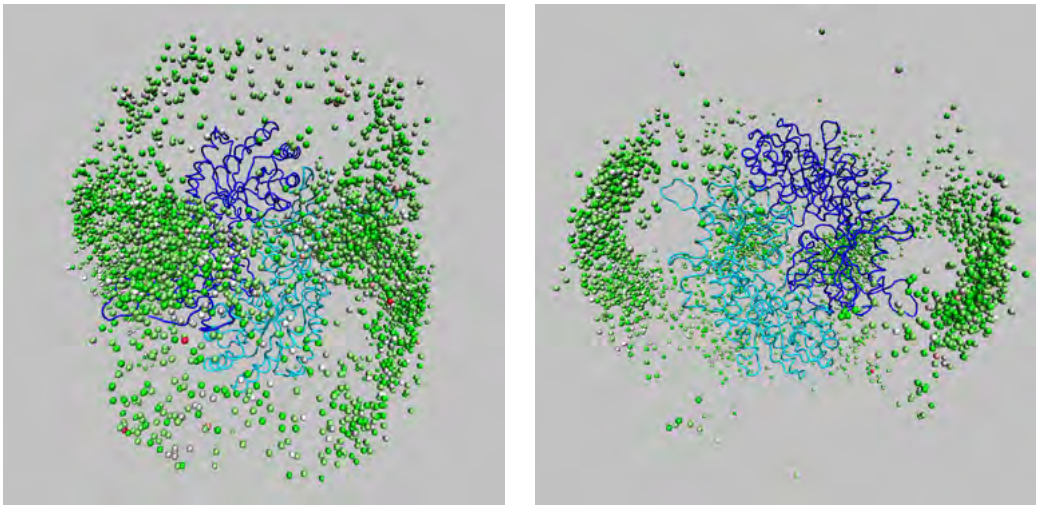


Figure B.28: Distribution of top 100 RP scores for docking of human (red) and *P. yoelii* (yellow) AdoMetDC/ODC. The notch overlaps are indicated by the hatched bars. Two overlaps can be observed for human and one for *P. yoelii* (The gap is too small to be visible on this scale). The RP score is plotted on the x-axis against the AdoMetDC/ODC model (yellow) or structure (red) combinations.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

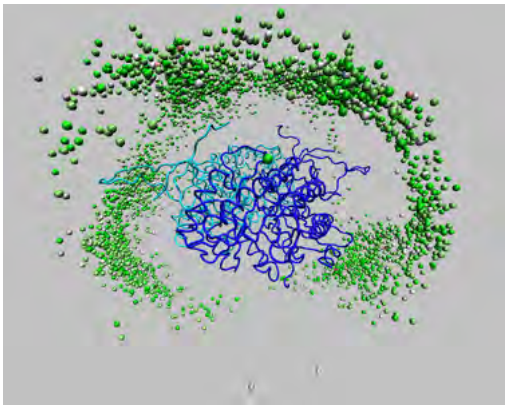
B.4 Distribution of centres of mass

B.4.1 AdoMetDC relative to ODC

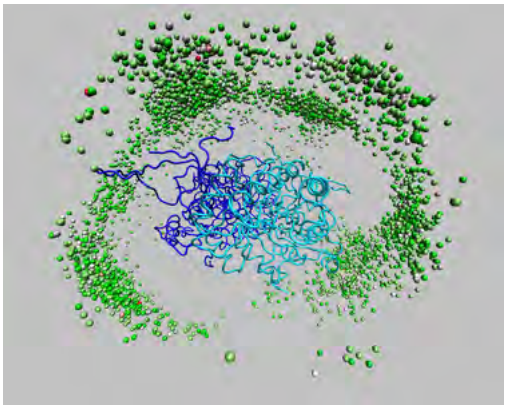


(a) Top

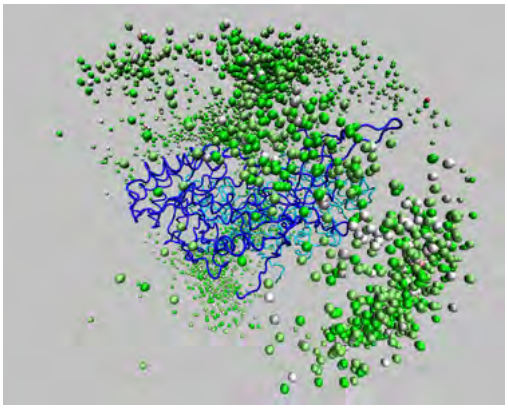
(b) Bottom



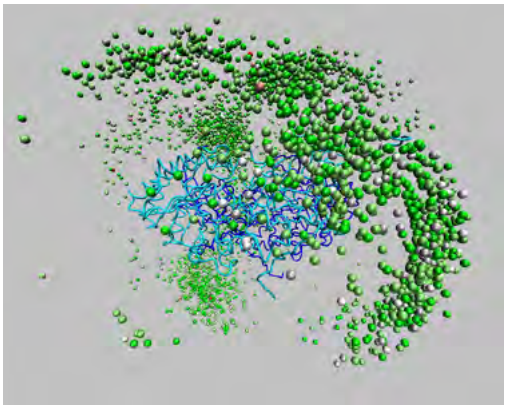
(c) Side 1



(d) Side 2

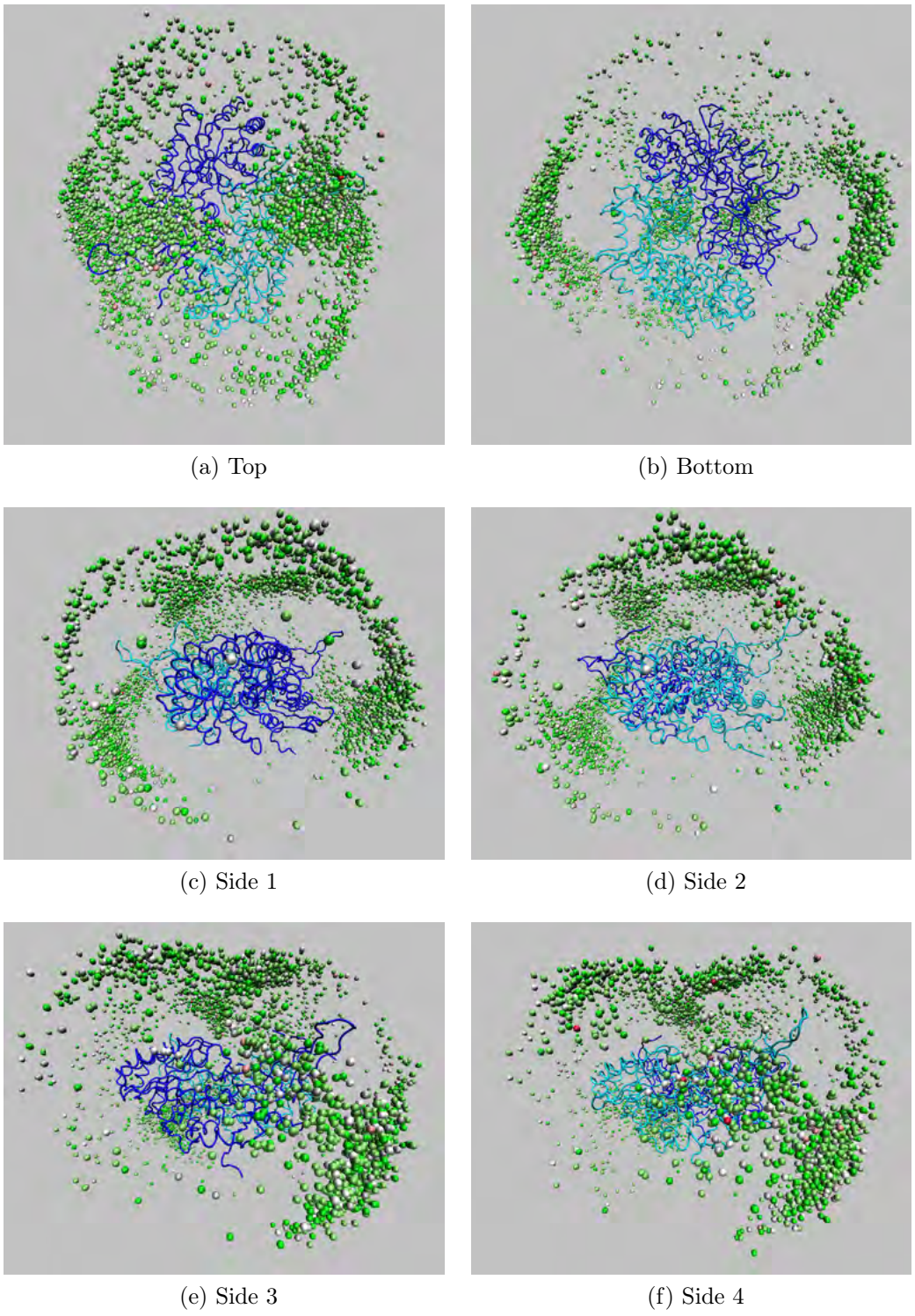


(e) Side 3



(f) Side 4

Figure B.29: Centre of mass (COM) distributions of *P. knowlesi* 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 (4.29: green → 11.53: red).

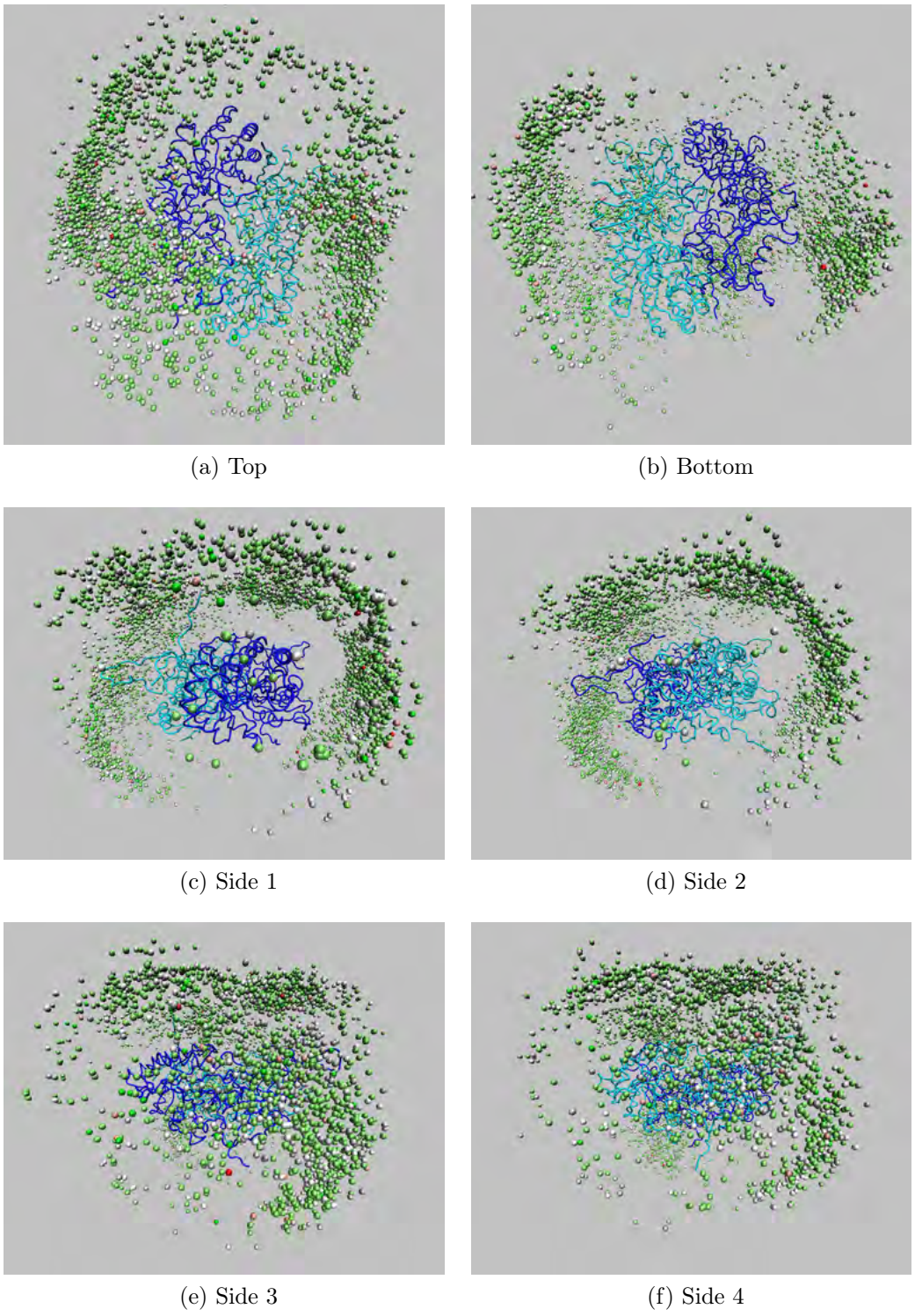


(a) Top (b) Bottom

(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

Figure B.30: Centre of mass (COM) distributions of *P. vivax* 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 (4.34: green → 11.42: red).



(a) Top (b) Bottom

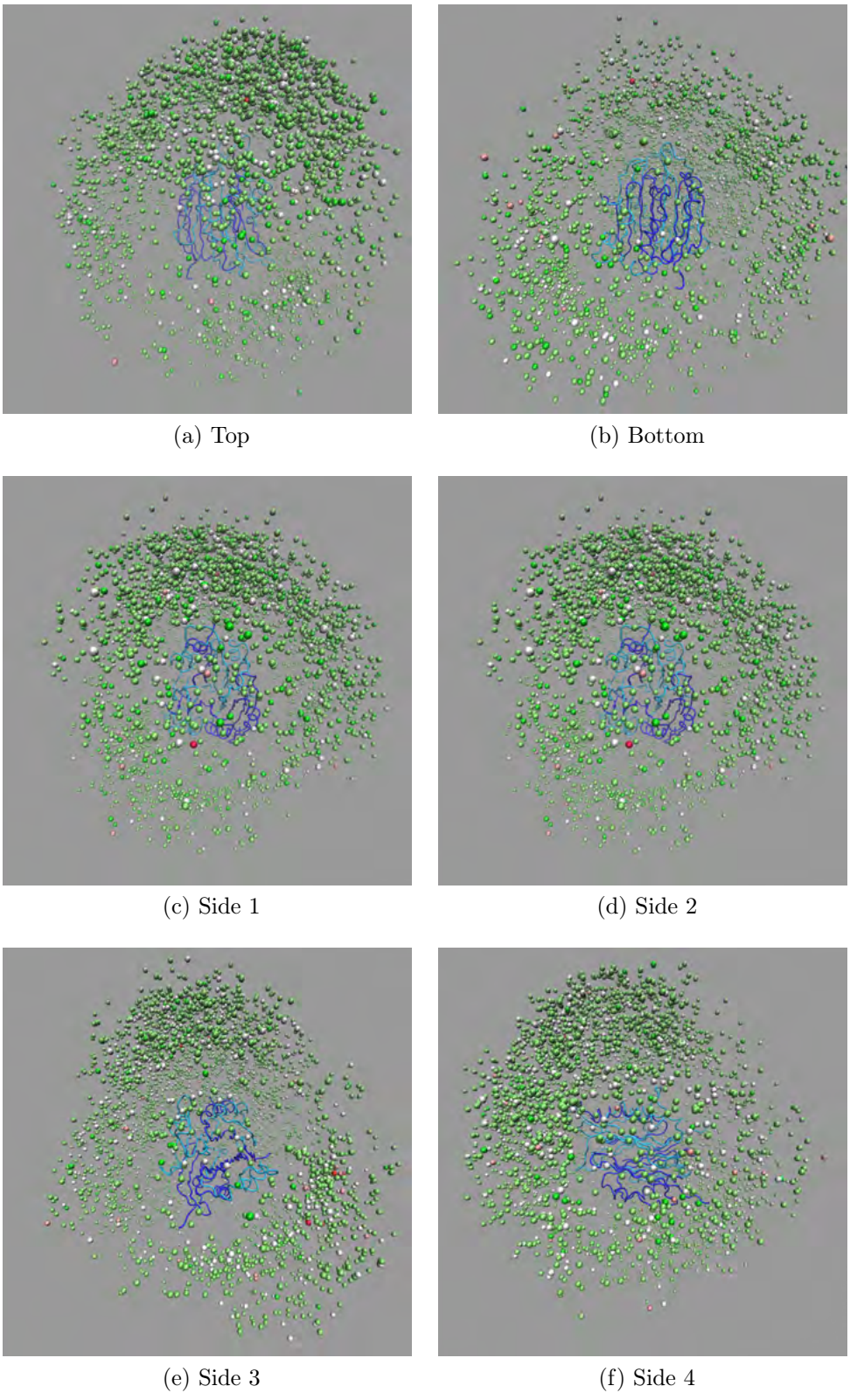
(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

Figure B.31: Centre of mass (COM) distributions of *P. yoelii* 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 (7.49: green → 12.28: red).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

B.4.2 ODC relative to AdoMetDC

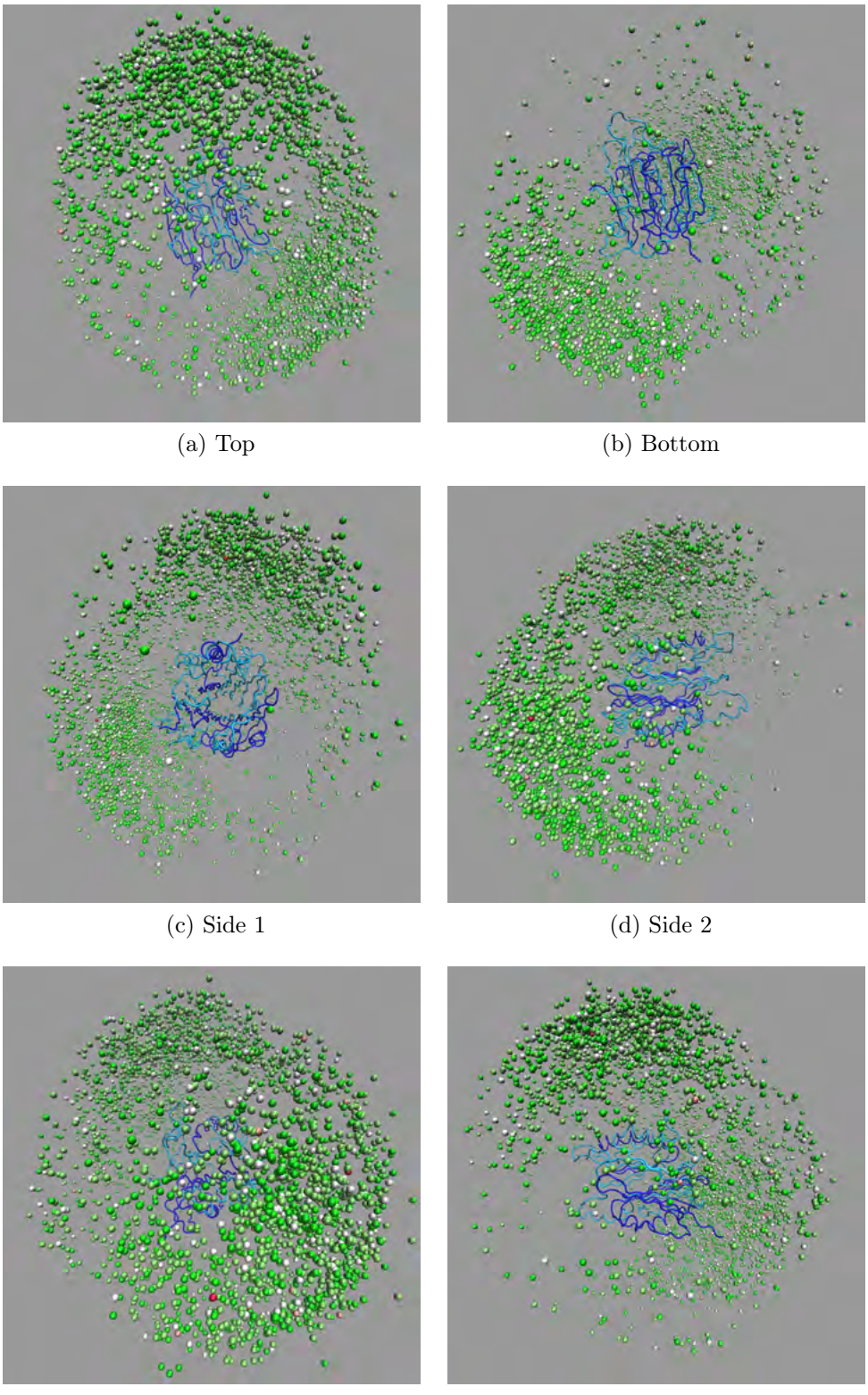


(a) Top (b) Bottom

(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

Figure B.32: Centre of mass (COM) distributions of *P. falciparum* 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 (4.87: green → 12.64: red).

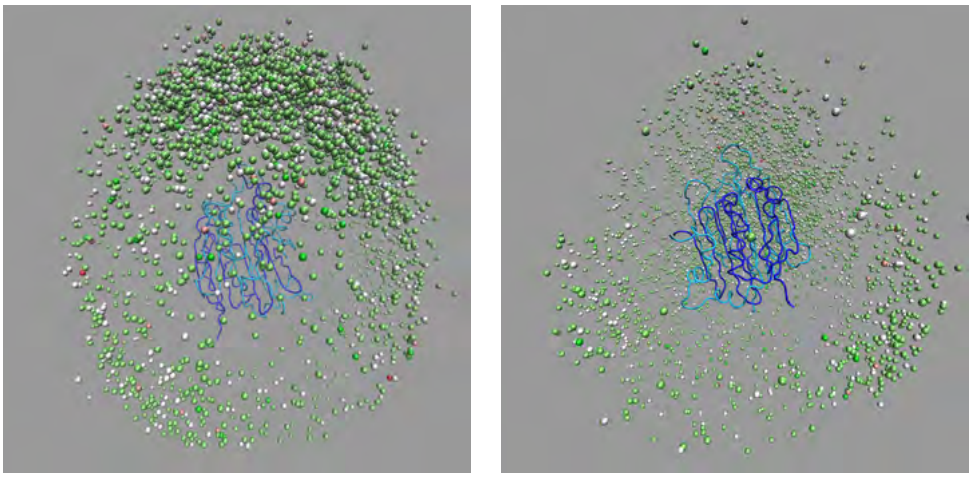


(a) Top (b) Bottom

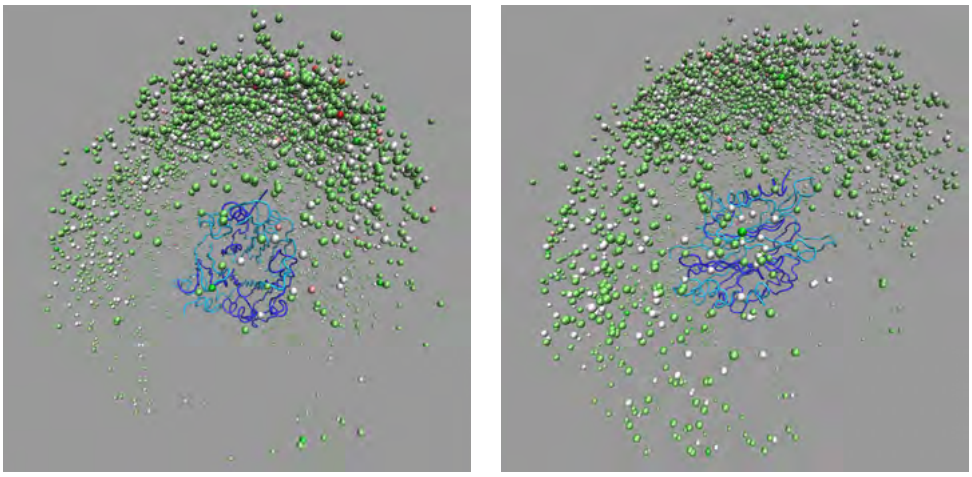
(c) Side 1 (d) Side 2

(e) Side 3 (f) Side 4

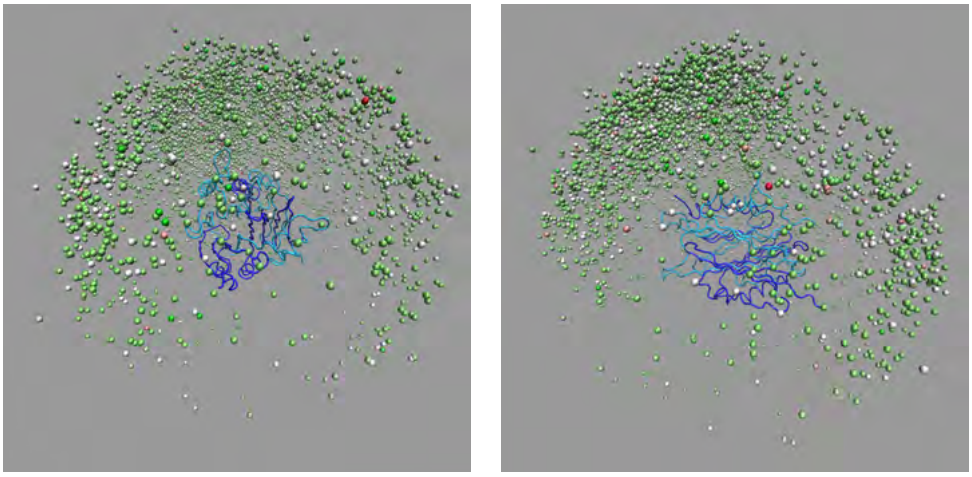
Figure B.33: Centre of mass (COM) distributions of *P. knowlesi* ODC relative to AdoMetDC. AdoMetDC chains A & C (blue) and (cyan) B & D are represented by C_α trace. The COMs of all top 100 (by RP score) from all dockings are represented as spheres (total 3600 positions). COM colour scaling is based on the RP score (4.29: green → 11.53: red).



(a) Top (b) Bottom



(c) Side 1 (d) Side 2



(e) Side 3 (f) Side 4

Figure B.34: Centre of mass (COM) distributions of *P. yoelii* 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 (4.89: green → 12.28: red).

B.5 Conserved interactions between AdoMetDC and ODC

B.5.0.1 All pairs

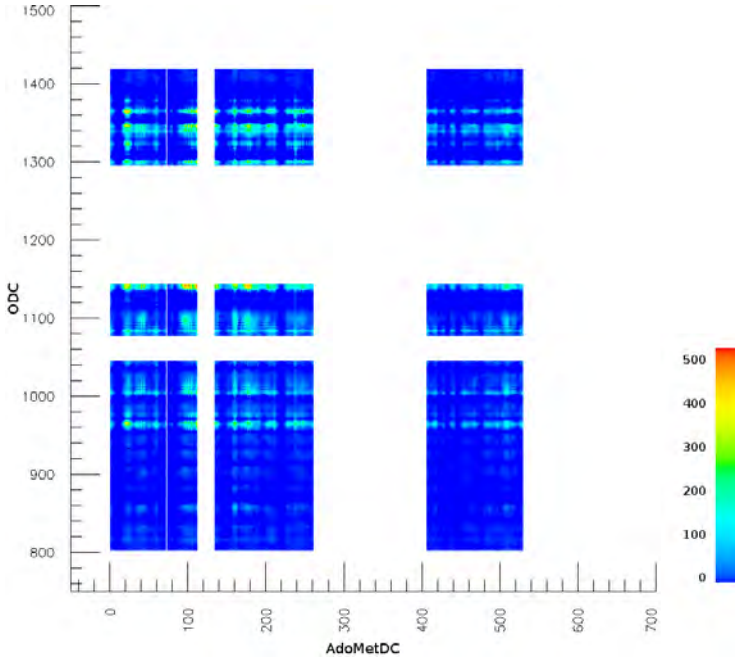


Figure B.35: Contact count heat-map for *P. falciparum*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_{α}^{ADC}-C_{α}^{ODC}<15 Å). The typical range is between 0-500, out of a possible 3600.

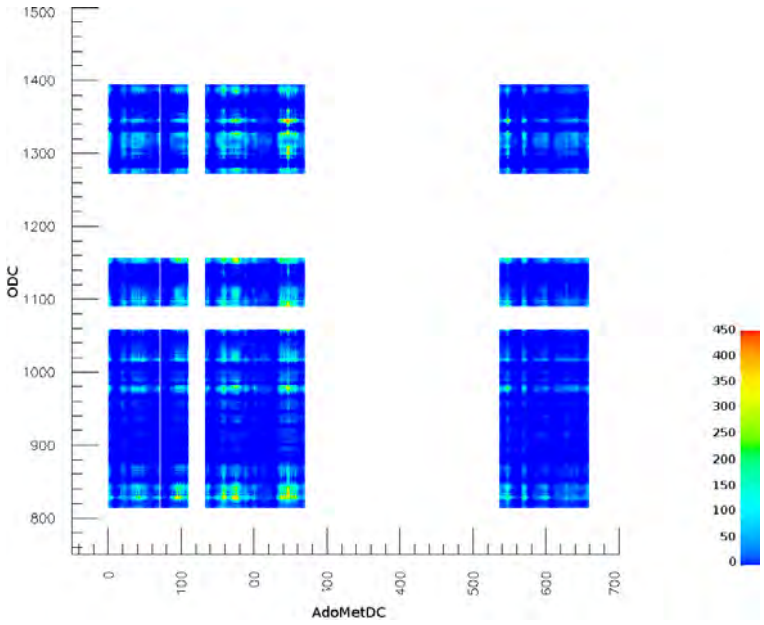


Figure B.36: Contact count heat-map for *P. knowlesi*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_α^{ADC}-C_α^{ODC}<15 Å). The typical range is between 0-500, out of a possible 3600.

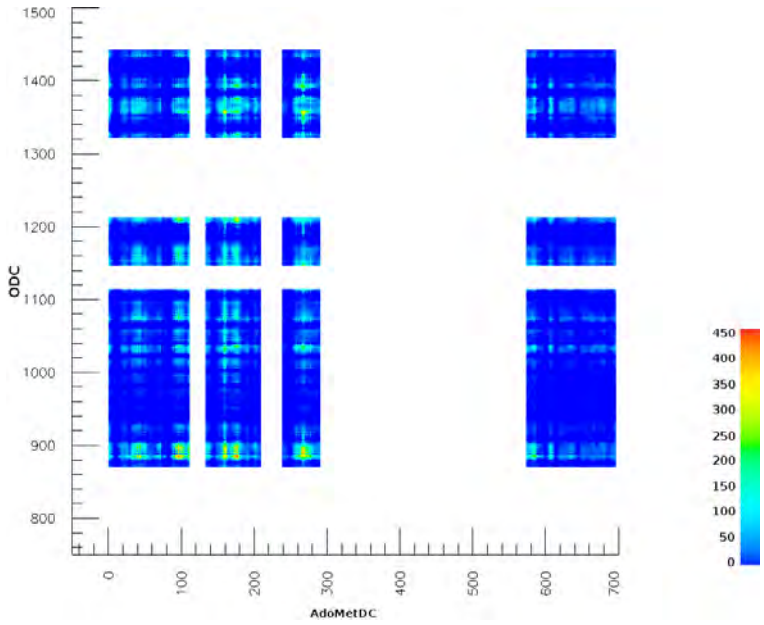


Figure B.37: Contact count heat-map for *P. vivax*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_α^{ADC}-C_α^{ODC}<15 Å). The typical range is between 0-500, out of a possible 3600.

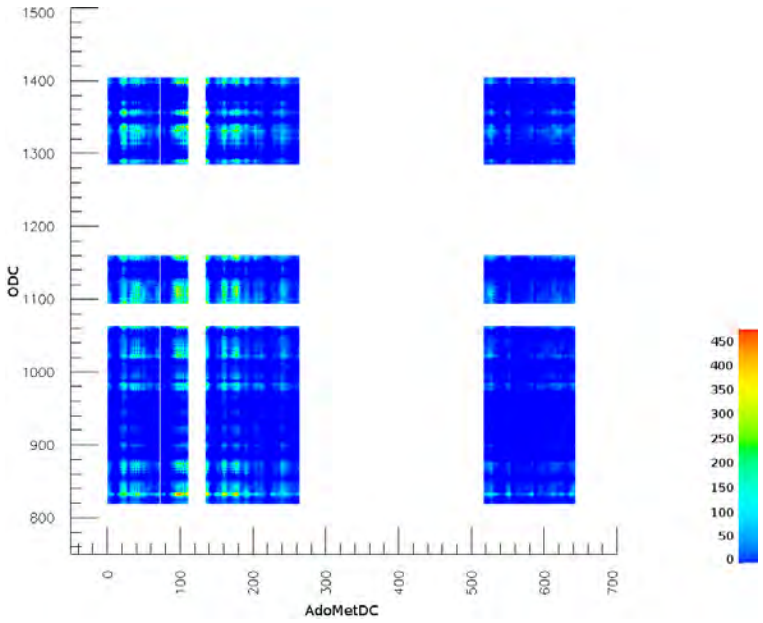


Figure B.38: Contact count heat-map for *P. yoelii*. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient (blue→green→red) corresponds to the number of times a residue pair makes contact (C_{α}^{ADC}-C_{α}^{ODC}<15 Å). The typical range is between 0-500, out of a possible 3600.

B.5.1 Conserved pairs

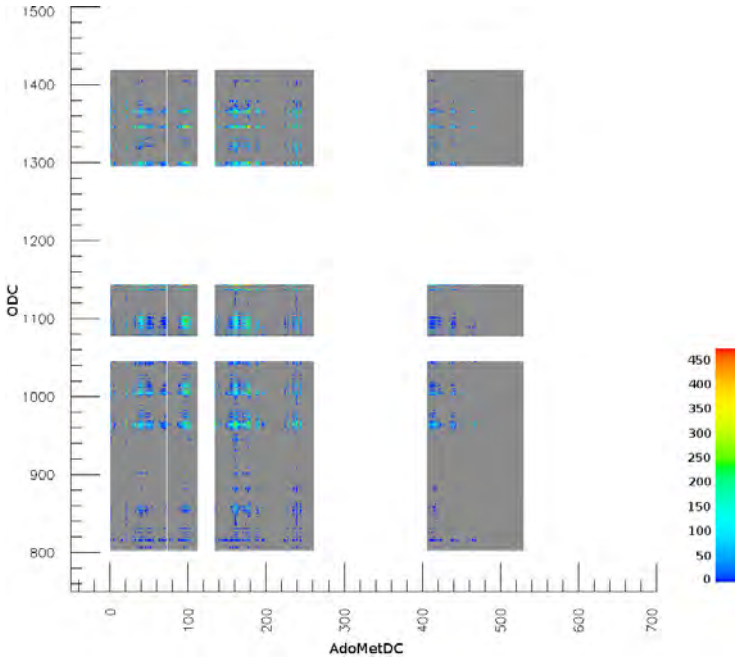


Figure B.39: *P. falciparum* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes 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.

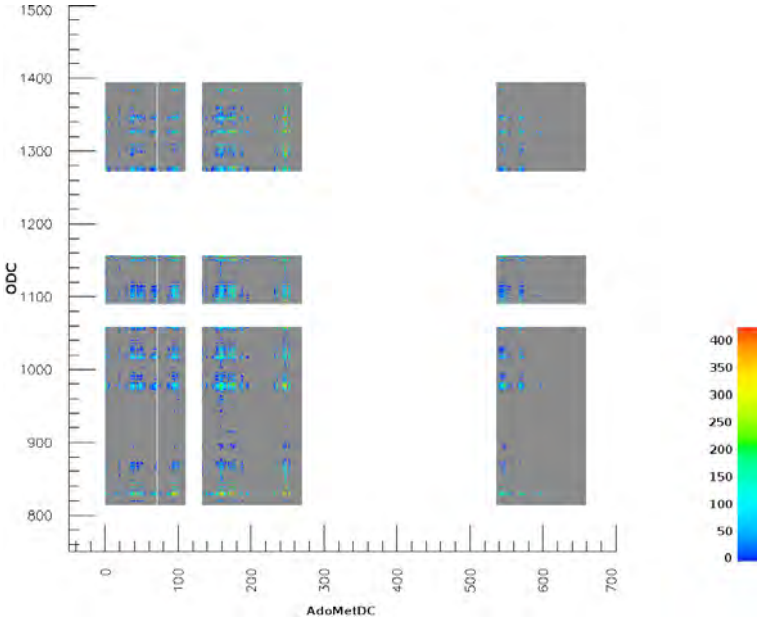


Figure B.40: *P. knowlesi* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes 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.

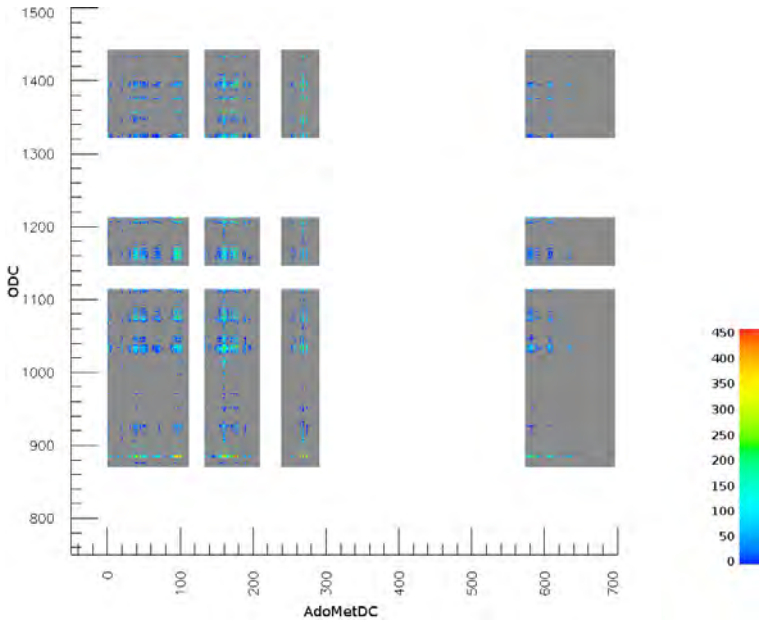


Figure B.41: *P. vivax* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes contact (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.

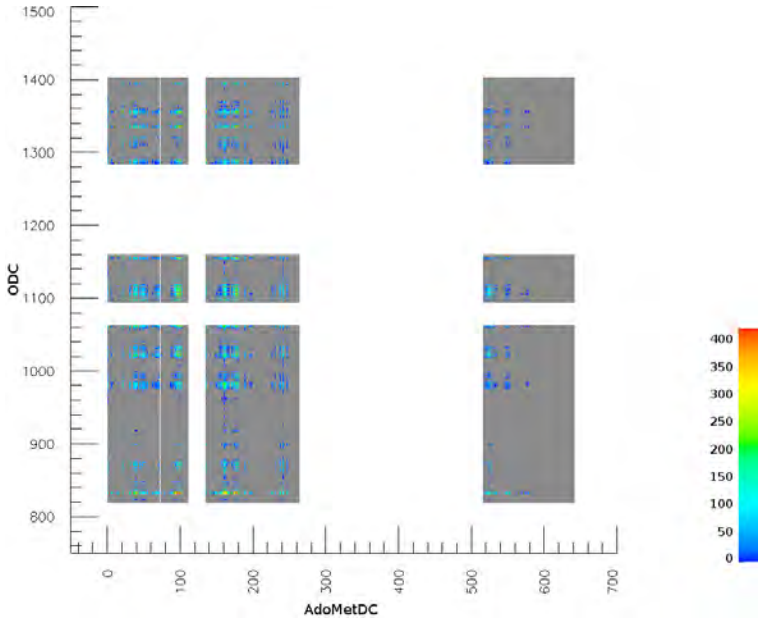


Figure B.42: *P. yoelii* contact count heat-map, restricted to pairs of identical residues across all *Plasmodium* species. AdoMetDC and ODC residue numbers are indicated on the x and y axes, respectively. The colour gradient corresponds (blue→green→red) to the number of times a residue pair makes contact (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.