In silico structural characterisation of *Plasmodium falciparum* dihydro-6-hydroxymethylpterin pyrophosphokinase dihydropteroate synthase (PPPK-DHPS)

by

Tjaart Andries Petrus de Beer

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae*

in the Faculty of Natural & Agricultural Sciences

Department of Biochemistry

University of Pretoria

Pretoria

February 2005

Acknowledgements

- My supervisors Prof AI Louw and Dr F Joubert at the University of Pretoria for all their generous support and guidance.
- The National Bioinformatics Network (NBN) for funding the visit of Prof. Lennart Nilsson from the Karolinska Institute, Sweden who presented the CHARMM course.
- Prof. L Nilsson for all his assistance in solving the problems we encountered.
- The National Research Foundation (NRF) and the University of Pretoria for awarding me the bursaries which enabled me to continue my studies.
- The International Society for Computational Biology (ISCB) for awarding me travel fellowships to attend the ISMB 2003 and 2004 conferences.
- My fellow students at the University of Pretoria for all their support and insightfull discussions.
- Developers of Open Source Software all over the world for providing top quality free software.
- To Lindie for all her love, patience and support.
- My parents for all their support and patience.
- Our Father in heaven.

Contents

List of Abbreviations

Å	Angstrom
ABNR	Adopted Basis Newton-Raphson
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
Ba	*Bacillus anthracis*
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CG	Conjugate Gradient
DHFR	Dihydrofolate reductase
DHFS	Dihydrofolate synthase
DHP	6-hydroxymethyl-7,8-dihydropterin pyrophosphate
DHPS	Dihydropteroate synthase
EC	Enzyme Commission
Ec	*Escherichia coli*
EM	Electron Microscopy
ESFF	Extensible Systematic Forcefield
FPGS	Tetrahydrofolylpolyglutamate synthase
GTP	Guanine triphosphate
Hi	*Haemophilus influenzae*
HP	6-hydroxymethyl-7,8-dihydropterin
HMM	Hidden Markov Model
K	Kelvin
kb	kilobases
kD	kilo Dalton
MANIC	6-methylamino-5-nitroisocytosine
MMFP	Miscellaneous Mean Field Potential
MSP	Merozoite Surface Protein

Mtb	*Mycobacterium tuberculosis*
NCBI	National Center for Biotechnology Information
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
ns	nanosecond
Pb	*Plasmodium berghei*
Pc	*Plasmodium chabaudi*
PDB	Protein Data Bank
Pf	*Plasmodium falciparum*
PPPK	Hydroxymethyldihydropteridine pyrophosphokinase
ps	picosecond
Pv	*Plasmodium vivax*
Py	*Plasmodium yoelli yoelli*
QM	Quantum mechanics
RESA	Ring-infected erythrocyte surface antigen
Sa	*Staphylococcus aureus*
SAN	Sulfanilamide
SD	Steepest Descent
TS	Thymidilate synthase

List of Figures

List of Tables

Typographical Conventions

- Amino acids will be referred to either by their one-letter or three-letter abbreviations. A mutation will be indicated by the amino acid being mutated followed by its number, followed by the new residue e.g. Ala437Gly or A437G indicates Alanine 437 being mutated to Glycine.
- Species will be referred to by a two or three letter abbreviation e.g. *Plasmodium falciparum* will be Pf and *Mycobacterium tuberculosis* will be Mtb.

Chapter 1

Introduction

1.1. The Malaria Parasite, *Plasmodium falciparum*

Malaria is caused by species of the *Plasmodium* genus. It infects humans, primates, rodents as well as birds. *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae* are the four main human parasites with the virulent *P. falciparum* implicated in the vast majority of human infections. The *Plasmodium* family belongs to a large, monophyletic group of organisms called apicomplexans. Their distinguishing feature is the presence of an apical organellar complex for invading the host cell. Apicomplexans include well known organisms such as *Cryptosporidium, Toxoplasma* and *Theileria* (Figure 1.1; Aravind *et al.*, 2003, Sherman 1998). There are some disagreements as to the origins of malaria. Most of the studies supports the idea that *P. falciparum* originated in Africa and then expanded to southeast Asia and South America but the date of this expansion varies from 10 000 years ago until a maximum of 400 000 years ago (Hartl, 2004). Different sequences tell different tales. Protein-coding genes suggest the occurrence of a rapid expansion (from a few hundred individuals) of the *P. falciparum* population 24 500-57 500 (upper limit) years ago with a probable common ancestor about 10 000 years ago (Rich *et al.*, 1998). The same protein-coding genes used by Rich *et al.* (1998) were analysed using Coalescent theory, which reported a common ancestor about 300 000-400 000 years ago from a population of about a 100 000 parasites (Hughes *et al.*, 2001). The great difference between these two estimates was narrowed down to the age and accuracy of data sets which the two groups used (Hartl, 2004). Another approach used introns in genes on chromosome 2 and 3. The introns for chromosome 2 indicated a common ancestor anything from 9 500-61 000 years ago but a study of 204 coding and non-coding regions from chromosome 3 suggested a common ancestor 100 000-180 000 years ago (Hartl, 2004). In an effort to decrease the uncertainty in the estimates of the population expansion of *P. falciparum,* it was decided to investigate the mitochondrial genome of the parasite. The *P. falciparum* mitochondrial genome is 6kb in size and inherited, as in most mammals, only from the female parent. The results showed that a rapid African expansion occurred about 10 000 years ago but some lineages are more ancient (50 000-100 000 years)

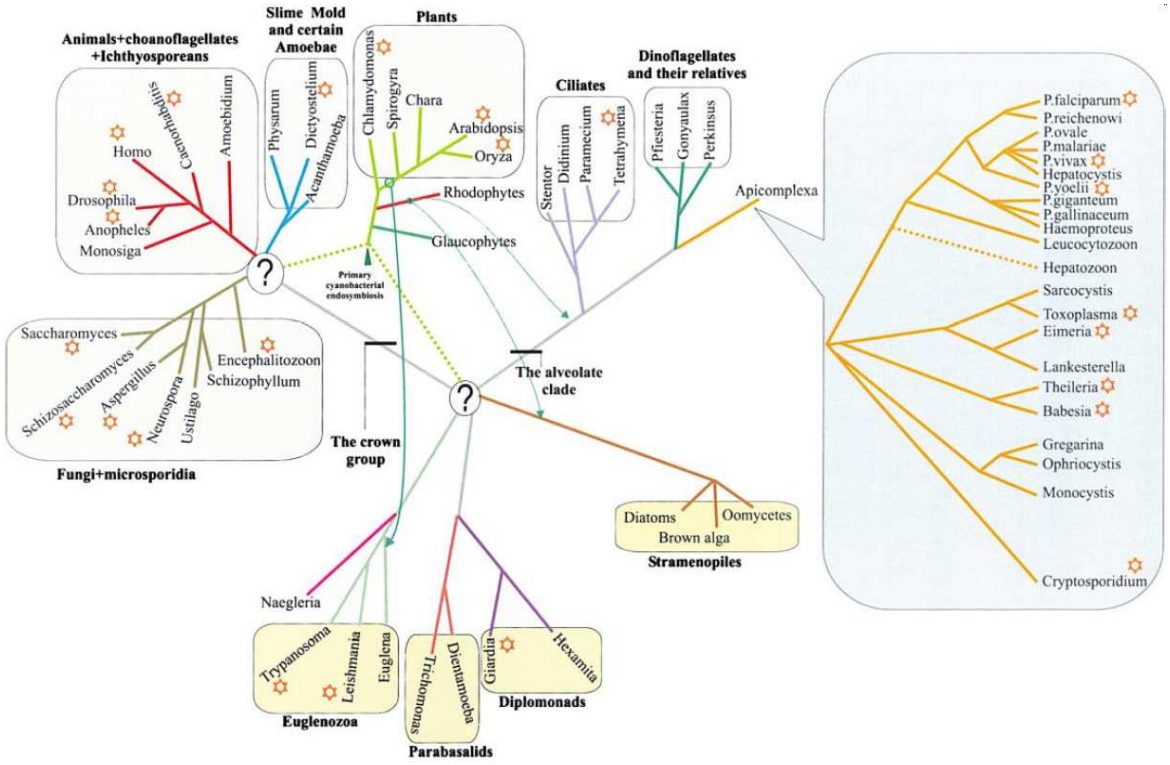


Figure 1.1: The phylogenetic context of *P. falciparum*. Question marks indicate areas where phylogentic relationships are not sorted out yet, red stars indicate organisms for which a complete genome sequence are either finished or in the final stages of processing. The thin arrows indicate secondary endosymbiotic events such as the acquisition of the chloroplast. Adapted from Aravind *et al.*, 2003.

that spread to South America and southeast Asia before the African expansion (Joy *et al.*, 2003). This data goes against the modern belief that the slave trade and Europeans introduced malaria into South America (Sherman 1998). Most of the studies on the population and genetic origins of *P. falciparum* supports the idea that a rapid expansion of *P. falciparum* occurred in Africa about 10 000 years ago.

The first historical evidence of malaria comes from Egyptian mummies and documents. Malaria symptoms were reported from Egypt (1570 B.C.), Persia (2000 B.C.), China (2700 B.C.) and by the Greeks (Homer 750 B.C.; Aristotle 384-322 B.C.; Plato 428-347 B.C.). The disease came to prominence during the Roman era and its name hails from the Italian name for "Roman fever" *mal'aria* ("bad air"). After that the disease spread slowly but surely across most of Europe and into Russia. By the early 1800's malaria had a worldwide distribution (Sherman, 1998). In the 1880's Charles L. A. Lavernan, Camillo Golgi and William G. MacCullum proved that malaria was caused by a single-celled organism which infects the erythrocytes of the human host. The end of the 1890's saw Ronald Ross and Patrick Manson prove that malaria was transmitted by the mosquito. Ross received the Nobel prize in 1902 and Lavernan received the Nobel prize in 1907 for their work on malaria (Sherman, 1998).

Plasmodium has a life cycle which is divided into three parts (Figure 1.2). Sporozoites infect the liver (the exoerythrocytic stage) where they develop into merozoites, which spread to the blood (erythrocytic stage). The merozoites develop into trophozoites, which develop into schizonts. The schizont develops

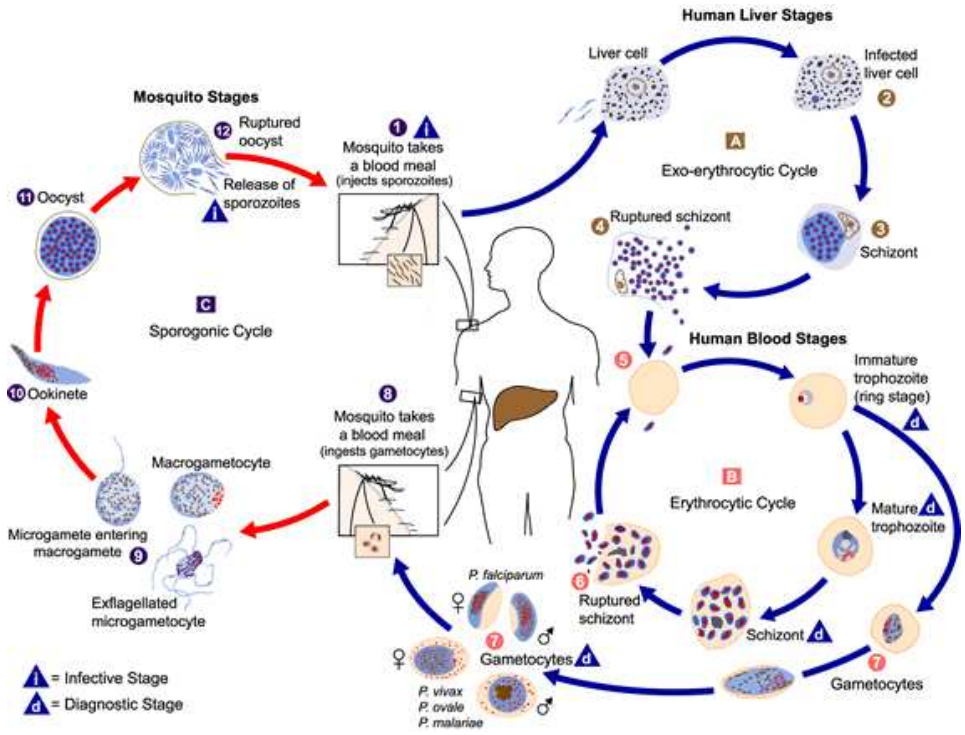


Figure 1.2: The life cycle of the malarial parasite. Source: http://www.well.ox.ac.uk/ich/images.

into trophozoites, which develop into male and female gametocytes. The gametocytes then need the mosquito to complete its life cycle by forming a zygote, which develops into sporozoites in the salivary gland of the mosquito (mosquito stage). In the case of *P. falciparum* the parasite is mostly carried by the *Anopheles gambiae* mosquito (Sherman, 1998).

Malaria is one of the most devastating diseases of the modern world. The worst affected region is sub-Saharan Africa with the worst affected countries suffering a 1.3% economic "growth penalty" each year as well as sustaining annual economic losses upwards of US$2 billion (Kilama, 2003; Figure 1.3). Each year malaria infects between 300 and 500 million people resulting in more than 1.5 million deaths, most of whom are children under the age of five (WHO Malaria fact sheet #94, http://www.who.int/mediacentre /factsheets/fs094/en/). Most of the fragile African economies cannot carry the economic and social burden of treating malaria. Alternative options to continous treatment are being considered such as vaccines, improving sanitation and living conditions and investigating low-cost, preventative methods such as bednets and raised houses (Kilama, 2003; Charlwood *et al.*, 2003). The current status of malaria vaccines will be dicussed in section 1.2.

The impact of malaria on the African economy is devastating. Although a major part of the problem is the cost of treating infections, the disease also results in a loss of income as adults are unable to work. This in turn forces adolescents to go searching for work to earn money and when infected, the same fate befalls them. The never ending cycle of infection, treatment and recovery drains the economy and hampers any attempts at economic recovery (Samba, 2004). The infection of children results in

young, female adolescents having to take care of the children. This reduces the earning capacity of the young adolescents as well as preventing the children from attending school. A strategy targeting young adolescents has been lacking, especially young females prior to conception and during and after pregnancy (Brabin *et al.,* 2005).

Some prevention strategies have been introduced but have met with limited success. The most notable was the introduction of insecticide-treated bednets. The treated bednets reduced infection of infants by 30%. The main problem with this approach is that the insecticide only lasts a few months and local communities do not have enough financial resources to obtain new insecticide (Snow *et al.,* 1997; Friedman *et al.,* 2003).

An alternative strategy is to target the malaria carrier, *Anopheles gambiae.* This type of vector control has met with limited success as *A. gambiae* has developed resistance against permethrin, the most common insecticide used (Ranson *et al.,* 2004). Dichloro-diphenyl-trichloroethane (DDT) spraying was the method of choice for years but due to the accumulation of DDT in animals its use was banned in most countries. Recently the effect of pyrethroid and organophosphate mixtures targeting *A. gambiae* proved to be feasible and may be the way forward in vector control against permethrin-resistant *A. gambiae* (Bonnet *et al.,* 2004). Vector control through limited contact between humans and parasites are hampered by the occurrence of vast, humid tropical areas in Africa.

Another problem in preventing the spread of malaria is the migration of people between areas. More and more land is being explored and opened up for economical exploitation as mines or for agriculture. This causes an influx and migration of workers resulting in malaria being carried to new areas by travellers and workers (Cruz Marques, 1987). The flux of workers between areas makes it more difficult to control the vector.

A sometimes overlooked factor in resistance is the occurrence of non-compliance with malaria treatment. This is especially prevalent where longer treatment periods (up to a week) are involved. Pinto *et al.* (2003) showed that a single 600mg dose of chloroquine is as effective as 600mg followed by two daily 450mg doses in the treatment of uncomplicated *P. vivax.* A single dose treatment will increase compliance in comparison with multiple dose treatments. The next section will focus on drug resistance in malaria and the way forward.

1.2. Drug Resistance in Malaria

The first drug used for malaria was quinine, which was isolated in 1820 from the bark of Cinchina trees in South America. Chloroquine was first synthesized in Germany and introduced at the end of World War II as a potent antimalarial. It remained the drug of choice for 40 years (Sherman, 1998). When resistance emerged after only ten years of use, the development of drugs such as mefloquine,

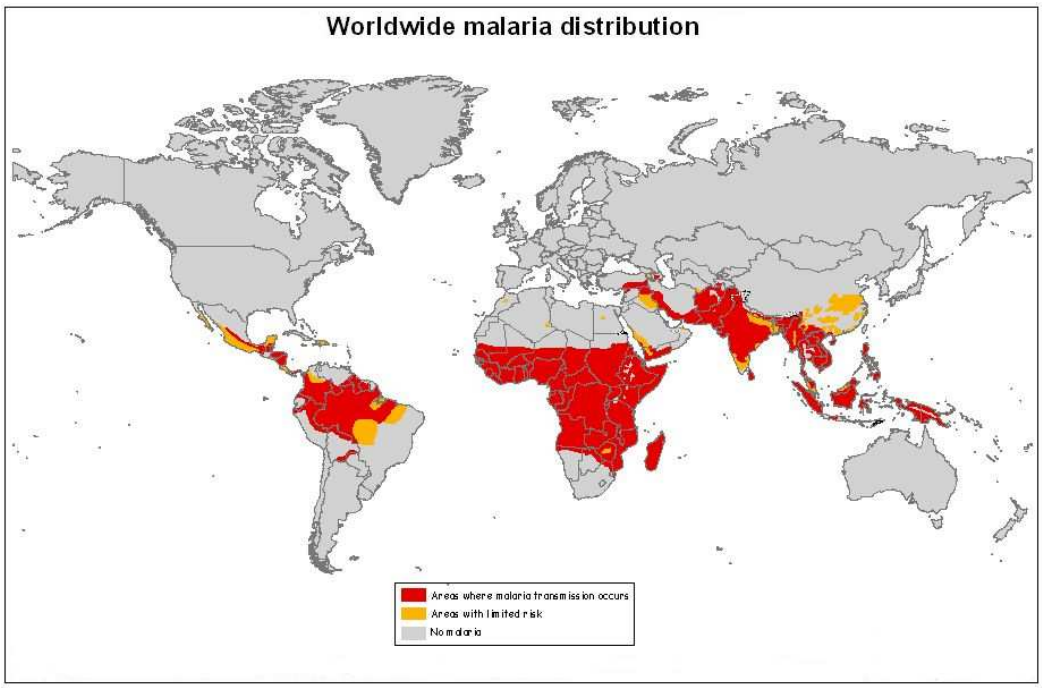


Figure 1.3: The distribution of malaria in 2002. Red shading indicates areas of malaria transmission, yellow shading indicates areas with a limited risk. Source: http://www.med.sc.edu:85/parasitology/blood-proto.htm.

sulfadoxine-pyrimethamine and artemisin derivatives were initiated. When resistance to some of the new drugs emerged, it signalled the beginning of multidrug-resistant *P. falciparum* (Wongsrichanalai *et al.,* 2002). As soon as a new drug was introduced, resistance developed within 10-20 years or in some cases even faster (Kremsner *et al.,* 2004; Table 1.1).

The emergence of multidrug-resistant parasites has complicated the treatment of malaria (Figure 1.4). Chloroquine resistance is now common in South America, Africa and southeast Asia (Sherman 1998). Sulfadoxine-pyrimethamine resistant parasites now occur in Gabon (Mawili-Mboumba *et al.,* 2001), Tanzania (Mutabingwa *et al.,* 2001), Kenya (Nzila *et al.,* 2000), Vietnam (Masimirembwa *et al.,* 1999), Brazil (Vasconcelos *et al.,* 2000), South Africa (Roper *et al.,* 2003) the Middle East (Wang *et al.,* 1997) and Thailand (Wongsrichanalai *et al.,* 2002). Mefloquine resistance occurs in Thailand and Africa (Wongsrichanalai *et al.,* 2002). Decreased quinine sensitivity occurs in South America and Africa, artesunate resistance in Thailand and dihydroartemisinin resistance in Africa (Wongsrichanalai *et al.,* 2002).

The cause of sulfadoxine-pyrimethamine resistance has been elucidated and is due to point mutations in the dihydropteroate synthase (DHPS) and DHFR enzymes, respectively (Triglia *et al.,* 1997; Yuvaniyama *et al.*, 2003). In DHFR up to four mutations confer resistance against pyrimethamine. These mutations include Ala16Val, Asn51Ile, Cys59Arg, Ser108Asn and Ile164Leu. Combinations of these mutations occur widely as well as in combination with mutations in DHPS. In DHPS there are five known resistance causing mutations which confer resistance to sulfadoxine. These are Ser436Ala/Phe,

Table 1.1: Drug introduction and resistance in malaria. Note the short times between the first introduction of the drug and emergence of resistance. Adapted from Wongsrichanalai *et al.*, 2002.

Antimalarial drug	Introduced	First reported resistance
Quinine	1632	1910
Chloroquine	1945	1957
Proguanil	1948	1949
Sulfadoxine-pyrimethamine	1967	1967
Mefloquine	1977	1977
Atovaquone	1996	1996

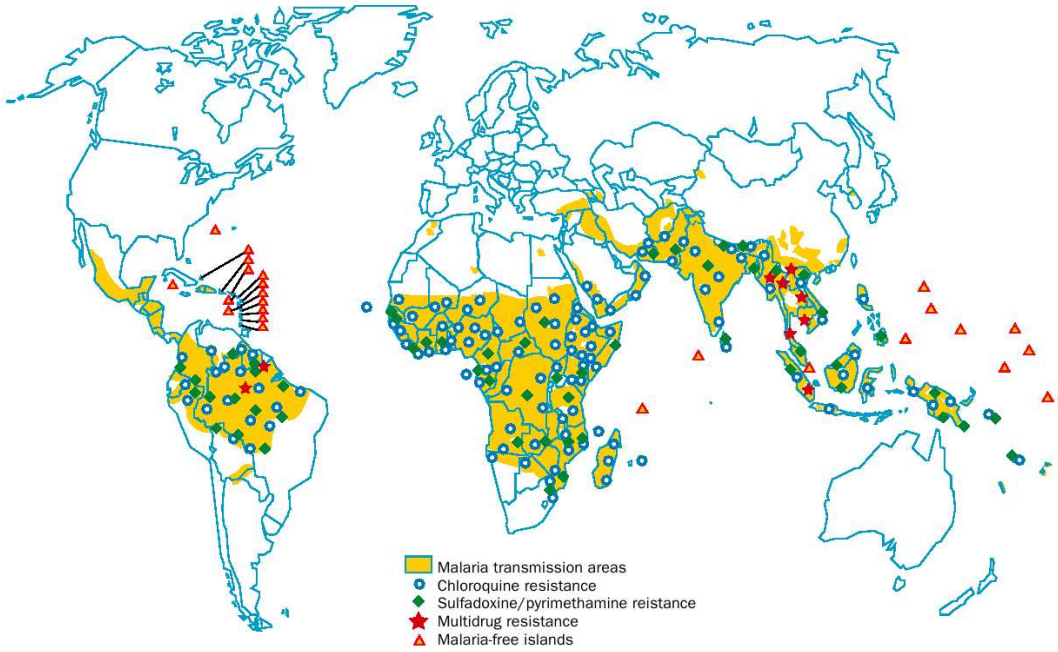


Figure 1.4: The occurrence of drug-resistant *P. falciparum*. Adapted from Wongsrichanalai *et al.*, 2002.

Ala437Gly, Lys540Glu, Ala581Gly and Ala613Ser/Thr. Up to four mutations have been observed in one genotype and these can combine with up to four mutations in the DHFR gene. This makes sulfadoxine-pyrimethamine treatment virtually useless in resistant genotypes (Wongsrichanalai *et al.*, 2002).

Triglia *et al.* (1994) established that gene amplification was not responsible for sulfadoxine resistance. Recently Nirmalan *et al.* (2004) showed that the antifolate drug target dihydrofolate reductase-thymidilate synthase (DHFR-TS) expression is upregulated in parasites when put under selective pressure by challenging the TS portion of DHFR-TS with fluoro-substituted bases. A factor which prevents antifolates from being a dependable treatment, is the occurrence of folate salvage and transport from the host by the *P. falciparum* parasite. Wang *et al.* (2004) showed that *P. falciparum* parasites have the ability to salvage pre-formed folate from the host or from growth medium. This reduces the effectiveness of

any antifolate drug as the folate pathway is circumvented. An ideal antifolate drug would block vital folate-metabolizing enzymes as well as block the transport of folate by the parasite.

The wide occurrence of resistance against the classic folate metabolism targets spurred investigations into alternative malaria targets. The ultimate solution to the malaria pandemic would be the development of a vaccine. This approach is being hampered by the lack of interest from industry and the lack of support for non-European affiliated countries. Current malaria vaccine targets include merozoite surface protein 1 and 3 and glutamine-rich protein (Kilama, 2003). Organisms that elicit immunity after a single infection are ideal for vaccine development. Unfortunately, in the case of malaria, partial immunity only develops after several years of exposure. This has been ascribed to antigenic polymorphism, antigenic switching and poor antigen immunogenicity. Polymorphisms in *P. falciparum* occur through at least three mechanisms, namely insertions/deletions of repeating units, point mutations and intragenic recombination. In MSP-1 polymorphisms have been attributed to the insertion of different numbers of a 9 bp tandem repeat into a region known as Block 2 in the gene. The insertion of these repeats coupled with intragenic recombination enables the parasite to continually change the antigenic epitopes and thus evade the host's immune system (Sakihama *et al.,* 2004; Ferreira *et al.,* 2003). Good (2005) suggested that a malaria vaccine should incorporate multiple antigens to elicit a maximal response. An alternative approach is using whole organisms to elicit an immune response (Good, 2005). One malarial vaccine used merozoite surface protein 1 and 2 (MSP-1 and MSP-2) as antigens (Genton *et al.,* 2002) but this was improved by including ring-infected erythrocyte surface antigen (RESA), MSP-1 and MSP-2 (Genton *et al.,* 2003). The studies showed an increase in the humoral response against the three antigens and shows promise as a potential vaccine against malaria.

A recently proposed drug-target is the hexose transporter of *P. falciparum.* The intraerythrocytic stages of *P. falciparum* are completely dependent on a glucose supply from the host which is taken up by a facilitative hexose transporter. The blocking of this transporter by long-chain *O*-3-hexose derivatives in cultured medium, resulted in the death of *P. falciparum* parasites. These long-chain *O*-3-hexose derivatives represent a potential new broad spectrum drug against different *Plasmodium spp.* (Joët *et al.*, 2004). Mitamura *et al.* (2003) presented the lipid metabolism of *P. falciparum* as a potential new drug target. Various approaches have now verified the existence of several *de novo* lipid synthesising pathways (glycerophospholipids, fatty acids and isoprenoids) and their study should present a few new targets in the fight against malaria (Joët *et al.,* 2003). One of the most potent antimalarials is Artemisinin which targets the sarco/endoplasmic reticulum Ca^{2+}-ATPase of *P. falciparum.* Derivatives of Artemisinins could provide a potent new antimalarial (Eckstein-Ludwig *et al.,* 2003). Vaidya (2004) highlights the vital importance of the mitochondrial and plastid genomes in the metabolism of *P. falciparum.* The mitochondrion and the plastid are vital to the parasite as their interruption was shown to be fatal for

the parasite. By studying the processes regulated by the genes of the mitochodrion and plastid genomes, new drug targets could also be identified (Vaidya, 2004).

The next section will focus on folate metabolism and the antifolates as well as their targets in *P. falciparum*.

1.3. Folates

1.3.1. Folate metabolism

Various processes in the cell require different co-factors. One of the most important co-factors is folate. Folate and its derivatives perform various functions in the cell: (1) it donates a methyl group during the conversion of homocysteine to methionine, (2) accepts a methyl group during the conversion of serine to glycine, (3) donates a methylene group to assist in synthesis of thymidilate, (4) tetrahydrofolate accepts a formimino group during the production of glutamic acid from formiminoglutamic acid, (5) donates a methyl group during the synthesis of pyrimidines and (6) can generate formate (Gilman *et al.,* 2001). DNA replication also demands pyrimidines and purines. Folate metabolism can provide the pyrimidines needed for replication.

The human host cannot produce folate and thus it is a vital ingredient in human diets. For the parasite to survive and reproduce it needs folate. Folate precursors are produced in the parasite by combining guanine triphosphate (GTP), *p*-aminobenzoic acid and glutamate (Figure 1.5). Dihydropteroate synthase (DHPS) is involved in producing 7,8-dihydropteroate, which in turn is converted to dihydrofolate via dihydrofolate synthase (DHFS). Dihydrofolate is reduced to tetrahydrofolic acid via the dihydrofolate reductase (DHFR) enzyme. Tetrahydrofolic acid is then methylated by serine hydroxymethyl transferase. The resulting methyltetrahydrofolate is used as a methyl donor in the *de novo* synthesis of pyrimidines. Methyltetrahydrofolate transfers its methyl group to deoxyuracil monophosphate to form deoxythymine monophosphate via thymidylate synthase (TS) (Gilman *et al.,* 2001).

In the *Plasmodium* species DHPS is a bifunctional enzyme together with hydroxymethyldihydropteridine pyrophosphokinase (PPPK), the enzyme which produces the substrate for DHPS (Triglia *et al.,* 1994). DHFR and TS are also linked in a bifunctional enzyme complex (Inselburg *et al.,* 1988) as well as DHFS and tetrahydrofolylpolyglutamate synthase (FPGS) (Salcedo *et al.,* 2001).

As folate is a vital ingredient for survival, a shutdown of the pathway would lead to parasite death. Various drugs target the folate pathway and these will be discussed in the next section. One of the problems with targeting the folate metabolism in *P. falciparum* is the existence of a folate salvage

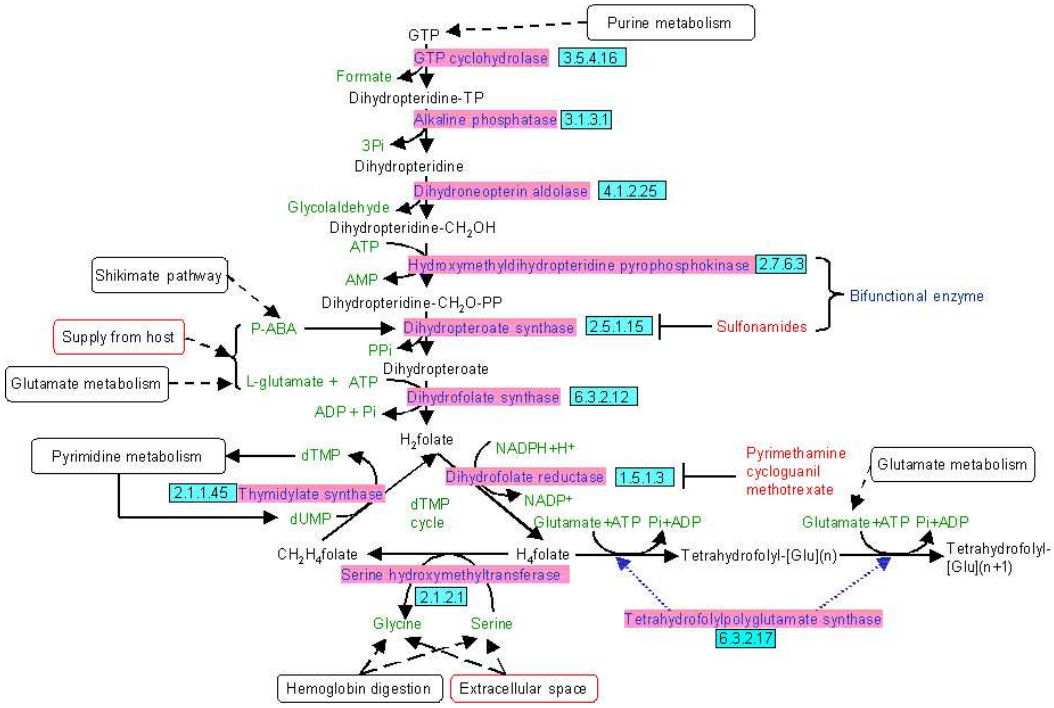


Figure 1.5: Folate metabolism in *P. falciparum*. Text in red indicates currently-used drugs and their targets in malaria. Source: http://sites.huji.ac.il/malaria.

pathway. This pathway can import folate from the medium around it (in the case of *P. falciparum*, from the erythrocyte) and thus, to a certain extent, circumvent the effect of anti-folates (Wang *et al.*, 2004).

1.3.2. Antifolates

Antifolates target certain enzymes in the folate pathway with the intention of interrupting the pathway and preventing the formation of vital compounds. Olliaro (2001) has divided the antifolates into two groups, Type-1 and Type-2. Type-1 antifolates include sulfone and sulfonamide drugs which mimic *p*-aminobenzoic acid. These drugs mainly target the DHPS enzyme and include drugs such as sulfadoxine and dapsone. Type-2 antifolates inhibit the DHFR enzyme and include drugs such as pyrimethamine, quinazolines and triazine metabolites (Olliaro 2001). The structures of the most widely used antifolates are shown in Figure 1.6.

Type-1 drugs block the synthesis of 7,8-dihydropteroate, a vital precursor in *de novo* folate synthesis, by mimicking *p*-aminobenzoic acid, a substrate of DHPS. Type-2 drugs prevent the reduction of H_2folate to H_4folate by DHFR and thus prevent the recycling of vital folate derivatives. Sulfadoxine, the most widely used antifolate type-1 drug, is usually given in combination with pyrimethamine (a type-2 drug) and is commercially known as Fansidar (Kasekarn *et al.*, 2004). One of the drawbacks of the sulfadoxine/pyrimethamine combination is the rapid emergence of resistance to both compounds. As mentioned in section 1.2 resistance to sulfadoxine and pyrimethamine is due to mutations in DHPS and DHFR,

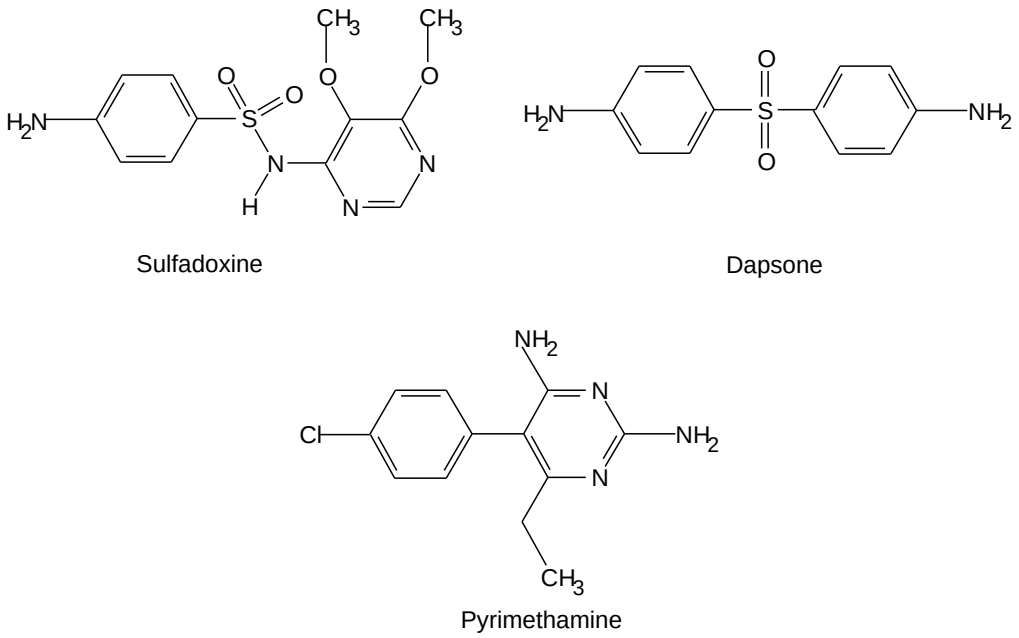


Figure 1.6: The most widely used antifolates in the treatment and prevention of malaria (Kasekarn *et al.*, 2004).

respectively. The next two sections will focus on the general structures, characteristics and mechanisms of action of PPPK and DHPS.

1.4. Dihydropteroate synthase (DHPS)

1.4.1. Structure of *P. falciparum* DHPS

The PfDHPS sequence was published in 1994 (Triglia *et al.*, 1994) and revealed that it was a bifunctional enzyme together with PPPK. The PfPPPK-DHPS enzyme complex consists of 706 amino acids with the PPPK enzyme (378 amino acids) occurring first in the sequence followed by the DHPS enzyme (328 amino acids). Triglia *et al.* (1994) localized PfPPPK-DHPS to a single copy on chromosome 8 of the *P. falciparum* genome. PfDHPS also contains an intron in the sequence (120bp in length), which contains conserved malarial splice acceptor and donor sites. PPPK-DHPS expression appears to happen throughout the life cycle of the parasite but peaks during the trophozoite stage. From immunoblots the size of PfPPPK-DHPS appears to be 68kDa (the size deduced from the amino acid sequence is 83kDa) but native enzyme isolation points to a size between 190-250 kDa, indicating that PfPPPK-DHPS may occur as a multimeric enzyme (Triglia *et al.*, 1994).

DHPS assumes the Triosephosphate Isomerase fold (TIM-barrel fold) in all known X-ray structures and thus it is realistic to expect that PfDHPS also occurs as a TIM-barrel. The next section will discuss the TIM-barrel fold.

1.4.1.1. The TIM-barrel fold

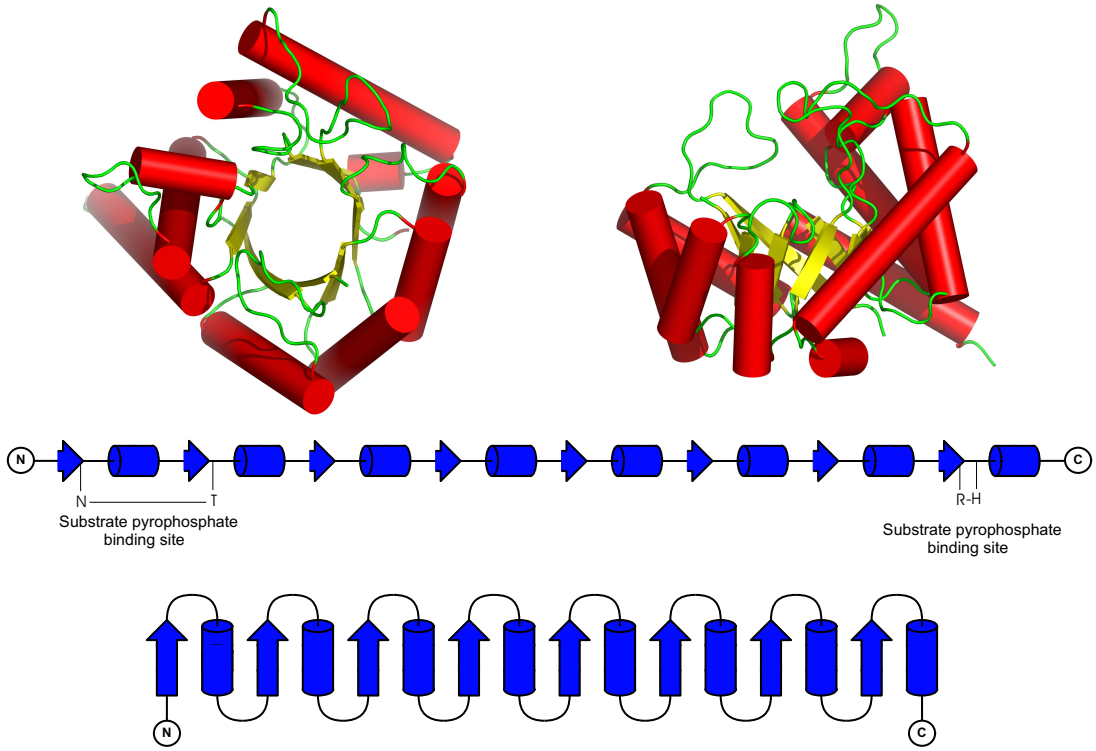


Figure 1.7: The structure of the TIM-barrel fold. Yellow indicates β-sheets, red indicates α-helices and green indicates the loops between the secondary structural elements. The middle figure shows a linear representation of the order in which the secondary structural elements follow one another and the bottom figure shows the orientation of the secondary structures in a TIM-barrel enzyme. Produced with PYMOL (DeLano, 2002; http://www.pymol.org) and TopDraw (Bond, 2003).

The TIM-barrel (β/α barrel) fold is one of the most common folds in the protein world and serves as a scaffold for various enzymatic functions. A search of the CATH database (Orengo *et al.*, 1997) revealed that at least 1870 proteins match the TIM-barrel topology (five of the six EC families are represented). Nagano *et al.* (2002) did an extensive review of TIM-barrel enzymes and found 21 homologous superfamilies and 76 families related by sequence. The TIM-barrel proteins almost always occur as enzymes (in contrast to structural proteins) and the co-factor/substrate usually includes a phosphate moiety. The active site always occurs at the C-terminal end of the barrel sheets and some members of the family exibit a structurally conserved phosphate binding site. The basic TIM-barrel topology (Figure 1.7) consists of 8 β-sheets and 8 α-helices, which fold into a barrel structure with the α-helices on the outside and the β-sheets on the inside (Nagano *et al.*, 2002). The barrels may also contain 7 or 9 β-sheets with secondary structural elements being numbered from the N-terminal end. TIM-barrels usually contain around 250 residues but may contain as little as 200 residues (Traut *et al.*, 2000). TIM-barrels also have a unique property in that the β-sheets usually exhibit a complete ring of hydrogen bonds (Wierenga, 2001). Wierenga (2001) also observed that the loops at the C-terminal end of the sheets are usually longer than those at the N-terminal end of the sheets and could thus contribute to the active site specificity.

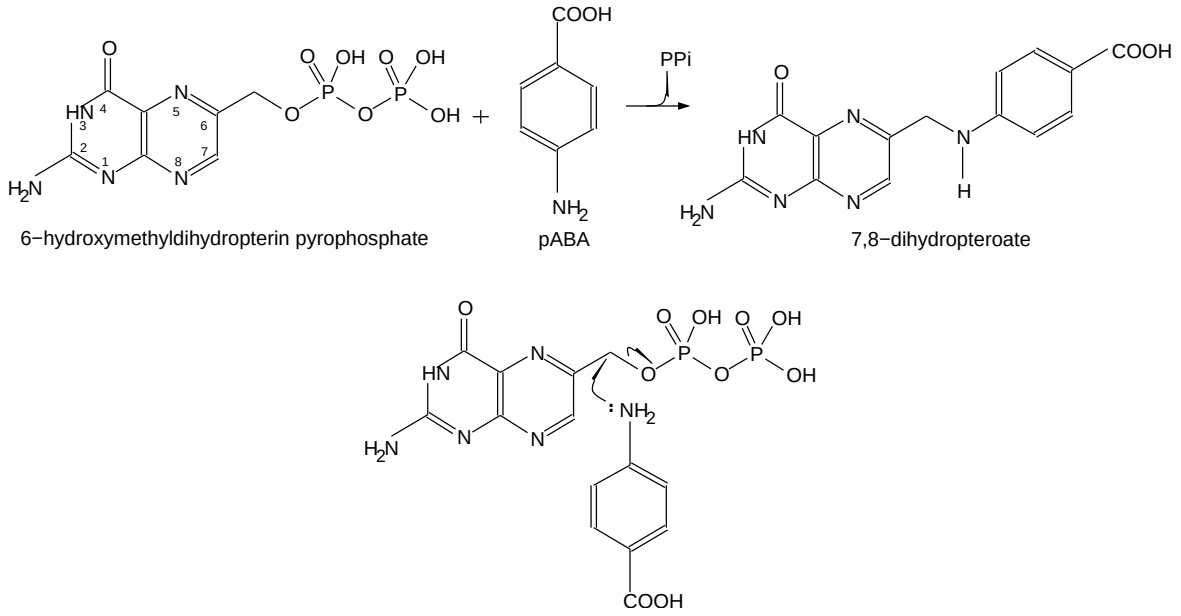


Figure 1.8: The reaction pathway of DHPS. Top: 6-hydroxymethyldihydropterin pyrophosphate (DHP) and *p*ABA condense to form 7,8-dihydropteroate with the production of one pyrophosphate group. Bottom: The mechanism of DHP and *p*ABA condensation (Baboaglu *et al.*, 2004).

The wide variety of TIM-barrel proteins led to the assumption that all TIM-barrels are evolutionary related. Lang *et al.* (2000) proposed that the original gene only coded for a half β/α barrel. This ancestral gene then underwent gene duplication and fusion to create a full β/α barrel, which underwent further gene duplication as well as functional and structural adaptations, resulting in a wide range of activities and functions. Copley *et al.* (2000) provides "statistically reliable evidence", which indicates that at least 12 of the 23 barrel superfamilies share a common origin. This data also supports the enzyme recruitment theory, which states that enzymes can be recruited between pathways, as well as supporting the idea that key metabolite-binding enzymes served as common ancestors for metabolic pathway evolution. The work done by Copley *et al.* (2000) also shows how one scaffold can adapt and fulfill various functions.

DHPS in *P. falciparum* exibits a conserved phosphate binding site and the substrate includes a phosphate group. This is consistent with the results for the DHPS family found by Nagano *et al.* (2002). The next section will discuss the function and mechanism behind *P. falciparum* DHPS.

1.4.2. Function and Mechanism of DHPS

The function of DHPS is to provide a precursor for the synthesis of folate (see section 1.3.1). Under the Enzyme Commission (EC) classification DHPS falls in subclass 2.5.1.15 (IUBMB 1992). This subclass of enzymes are involved in the transfer of alkyl or aryl groups, other than methyl groups. The mechanism of action of DHPS is based on a condensation reaction between *p*-aminobenzoic acid and 6-hydroxymethyldihydro-pterin pyrophosphate to form dihydropteroic acid (Figure 1.8) with a Mg^{2+} ion

as co-factor. The condensation reaction occurs through a SN_2 nucleophilic attack of the nitrogen on the carbon and the release of a pyrophosphate group. This reaction produces dihydropteroic acid which is used by dihydrofolate synthase to produce H_2 folate (Baboaglu *et al.*, 2004).

1.5. Hydroxymethyldihydropteridine pyrophosphokinase (PPPK)

1.5.1. Structure of *P. falciparum* PPPK

P. falciparum PPPK is part of the bifunctional PPPK-DHPS complex and contains one intron of 194 bp in length (Triglia *et al.*, 1994). PfPPPK consists of 378 amino acids and has two inserts when compared to other PPPK enzymes. Triglia *et al.* (1994) has shown that these inserts are not introns as cDNA amplification of the flanking areas correspond to the genomic DNA sequence. PPPK is connected to DHPS through a hinge region but this hinge region shows no homology to the only other known bifunctional PPPK-DHPS enzyme of *P. carinii* (Triglia *et al.*, 1994). From homology to other PPPK genes it is evident that PfPPPK folds into a Ferrodoxin-like fold (Figure 1.9; PPPK is also known as HPPK; Blaszczyk *et el.*, 2000; Hennig *et al.*, 1999; Stammers *et al.*, 1999). This will be discussed in the next section.

1.5.1.1. The Ferrodoxin-like fold

The Ferrodoxin-like fold falls in the alpha and beta class of proteins in SCOP (Hubbard *et al.*, 1999). It is described as an α/β sandwich with an antiparallel β-sheet. The Ferrodoxin-like fold consists of 5 α-helices and 4 β-sheets. The enzyme folds into a four stranded anti-parallel β-plate flanked on one side by 2 α-helices and on the other side by 3 α-helices. The secondary structures assumes a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4\alpha_3\alpha_4\alpha_5$ arrangement. The $\beta\alpha\beta\beta\alpha\beta$-motif is common to other nucleoside diphosphate kinases but PPPK contains an extra three α-helices (Stammers *et al.*, 1999). Nucleoside kinases usually bind ADP between the $\alpha_2 - \beta_4$ and the $\beta_2 - \beta_3$ interconnecting loops but PPPK binds ATP between the connecting loops of $\alpha_2 - \beta_4$ and the initial loop of the C-terminal region (Stammers *et al.*, 1999). The next section will describe the mechanism behind PPPK.

1.5.2. Function and Mechanism of PPPK

According to the Enzyme Commission (EC) classification, PPPK falls under subclass 2.7.6.3. This subclass is classified as diphosphotransferases (IUBMB 1992). PPPK catalyses the transfer of a pyrophosphate group from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP) to form DHP (Figure 1.10; Blaszczyk *et al.*, 2000). DHP is used by DHPS to form pteroic acid, a precursor to folate. The PPPK reaction

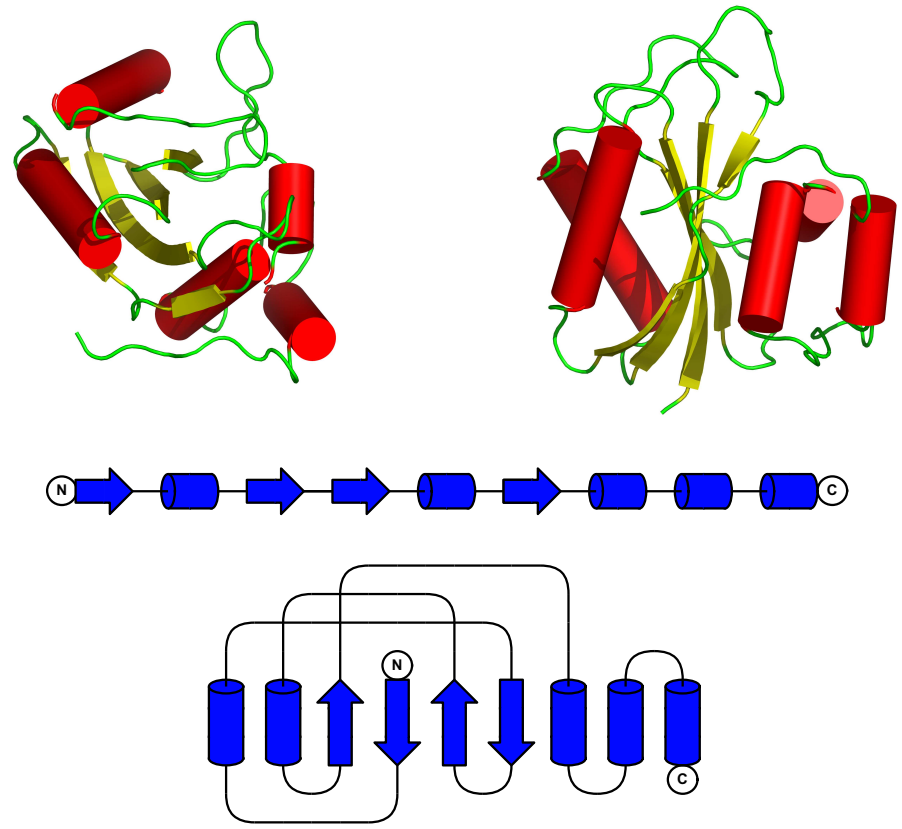


Figure 1.9: Top: The structure of the Ferrodoxin-like fold. The middle figure indicates the sequence in which the secondary elements follow one another. The bottom figure indicates the topology of the Ferrodoxin-like fold. Produced with PYMOL (DeLano 2002, http://www.pymol.org) and TopDraw (Bond 2003).

has two Mg^{2+} ions as co-factors, which helps to anchor the pyrophosphate groups during catalysis. The reaction occurs in a tightly closed enzyme with loops 1-3 closing over the active site. Loop 2 and 3 seems to close over the area where the two catalytic Mg^{2+} ions are located. In the closed structure the three loops associate with one another through a network of hydrogen bonds, which stabilises the interaction between the loops (Blaszczyk *et al.*, 2000).

1.6. Aims

The PfPPPK-DHPS complex contains one of the enzymes against which anti-folate resistance has developed. In order to investigate the mechanism behind sulfa-drug resistance a structural model of the complex is needed. Credible models should contain a reliable representation of the active sites and all the residues involved in resistance. Such models should aid explanations of the effect of resistance-causing mutations on substrate binding, drug binding and protein movement. To gain a more complete understanding of the PfPPPK-DHPS complex, the active sites as well as the interaction between the proteins and their natural substrates need to be investigated to reveal any possible similarity in ligand binding. To achieve these aims the following strategy was implemented:

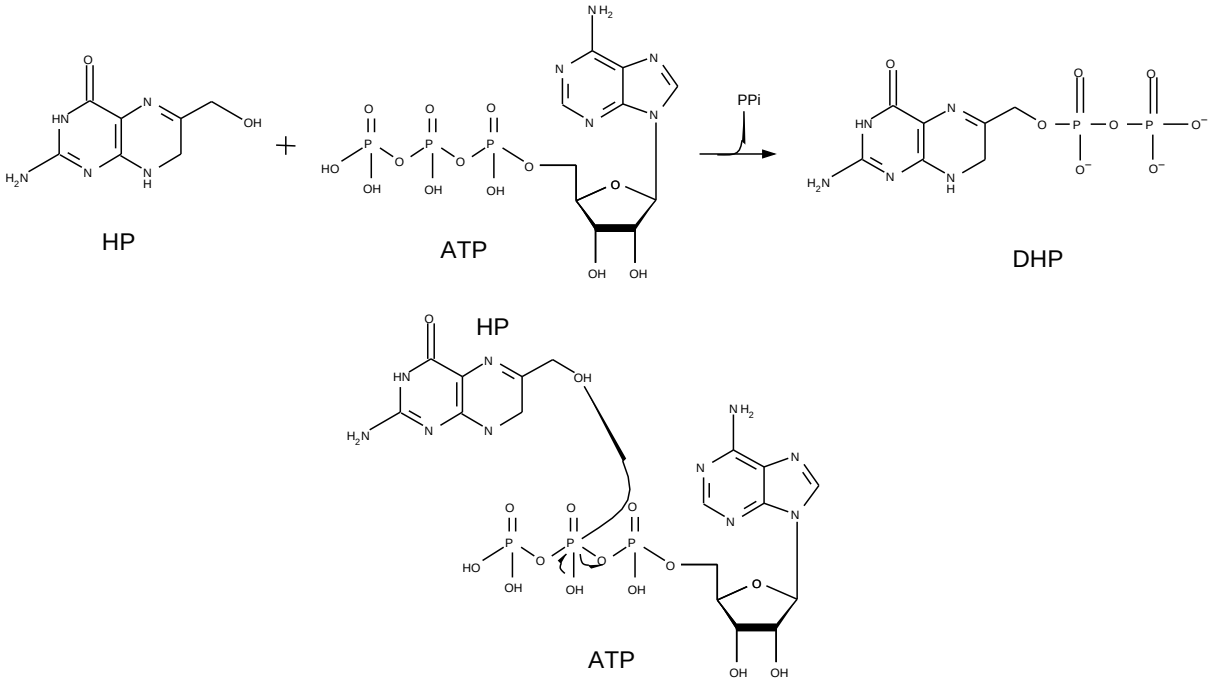


Figure 1.10: Top: The reaction pathway of PPPK. Bottom: The mechanism of PPi transfer from ATP to 6-hydroxymethyl-7,8-dihydropterin (HP) by PPPK as proposed by Blaszczyk *et al.* (2000).

- Homology models of *P. falciparum* PPPK and DHPS were constructed separately to investigate the structural characteristics of the two enzymes (**Chapter 2**).
- The natural substrates were docked into the active site of PfDHPS to investigate the interaction between PfDHPS and the substrates. The interactions between sulfa-drug resistant strains and the natural substrates were also investigated (**Chapter 2**).
- Sulfadoxine was docked into the active site of wild type DHPS, as well as the different resistant strains to investigate the mechanism behind sulfa-drug interaction in PfDHPS (**Chapter 2**).
- The effects of the mutations on the binding and movement of the substrates as well as on sulfa-drugs in the PfDHPS active site, were investigated with the use of Molecular Dynamics (**Chapter 3**).
- The extent of loop movement in PfDHPS and PfPPPK, as well as the effect of resistance-causing mutations on these loop movements, were investigated with Molecular Dynamics (**Chapter 3**).

Chapter 4 will provide a brief summary as well some concluding remarks.

Chapter 2

Structural Modelling of *P. falciparum* PPPK-DHPS

2.1. Introduction

To study the function of proteins it is vital to establish the coordinates of the important residues (preferably to atomic resolution) to gain insight into the mechanism of catalysis, substrate orientation and substrate binding. There are various methods available to obtain protein structure coordinates such as X-ray crystallography, cryo-electron microscopy and NMR measurements. These methods give various levels of resolution, with X-ray crystallography usually giving the highest resolution (down to 0.54Å (Jelsch *et al.*, 2000)) and cryo-EM the lowest (only down to secondary structure discrimination, 8-10Å). One of the Achilles heels of X-ray crystallography is the need to overexpress the protein or isolate sufficient quantities from natural sources. The latter method is a long and tedious process and thus overexpression in heterologous systems is the preferred method. Most proteins, especially malarial proteins, do not overexpress well in *E. coli* due to the rich A+T-codon bias (Aravind *et al.*, 2003). However, a protein may express in high enough yields but may not fold correctly and thus form inclusion bodies. Some proteins such as *P. falciparum* S-adenosyl-methionine decarboxylase, are expressed at low levels and are only stable in solution for a few hours, thus the risk increases for low, inactive protein yields (Birkholtz *et al.*, 2004). Once sufficient quantities of pure protein have been obtained it has to subjected to the procedure of crystallization which is regarded by most scientists as a "black box". Some effort has gone into trying to elucidate crystallization parameters, such as pI of the protein (Kantardjieff *et al.*, 2004) and the pH at which crystallization occurred (Dougall *et al.*, 2004), but they have met with limited success. In spite of these hurdles on the way to obtaining protein crystals, a resolution of 0.54Å is currently obtainable (Jelsch *et al.*, 2000).

NMR is one alternative option but suffers from some of the same problems as X-ray crystallography (expression levels, protein isolation). An advantage of NMR is that dynamic structures are obtained but currently only proteins up to ~80 kD can be investigated (Tugarinov *et al.*, 2002). In the early 70's researchers started to investigate computer models as an alternative to study protein structures.

The advantage of using computer modelling is the speed with which a protein can be modelled but the resolution achieved is dependent on the percentage identity between the target and template. The field of computer-aided protein modelling expanded during the 90's with the release of the MODELLER package (Sali *et al.*, 1993).

2.1.1. Homology modelling

Homology (or comparative) modelling is the process where the structure of a protein is modelled based on a closely related protein with a known structure. It is usually assumed that the proteins are homologous, meaning they have a common evolutionary ancestor. The 3D structures of proteins change little with evolution and thus the assumption which underpins comparative modelling is that proteins with similar sequences will have a similar structure (Marti-Renom *et al.,* 2000). The opposite is also true in that proteins with a similar structure do not necessarily share the same sequence (Marti-Renom *et al.,* 2000).

There are three main methods used to construct models: modelling by rigid-body assembly, modelling by segment matching and modelling by satisfaction of spatial restraints. Rigid-body assembly is the method whereby a protein model is constructed from small rigid pieces. These pieces are obtained from sequence alignments and are based on conserved core regions, loops and side-chain conformations. The backbone is constructed by taking the average C_α atom coordinates from the conserved part of the template structures. The rest of the main chain atoms are then superimposed from the template sequence with the highest similarity to the target. The loops are built after a library search for similar structures and the side-chains are built using a combination of template side-chain conformation and preferred side-chain conformation. The COMPOSER package (Sutcliffe *et al.*, 1987) applies this method of modelling.

Modelling by segment matching uses short peptide sequences to construct a model. By searching a database with the specific short sequence and matching C_α atom coordinates, short pieces of protein can be assembled. These short pieces are then put together and a conformational restraint is applied. The advantage of this method is that short insertions and deletions as well as side-chain atoms can be constructed. The method is implemented in the SEGMOD package (Levitt, 1992).

Spatial restraint satisfaction modelling is one of the most promising methods as various restraints can be used to guide protein construction. A program like MODELLER assumes that aligned residues will have similar restraints such as bond angles, dihedral angles and stereochemistry and uses these values to construct the model (Figure 2.1; Sali *et al.*, 1993).

The spatial restraints are then combined with the CHARMM22 forcefield parameters (to enforce proper stereochemistry) into an objective function (MacKerrell *et al.*, 1998). The model then undergoes optimization of the objective function in Cartesian space. The inclusion of the CHARMM22 forcefield

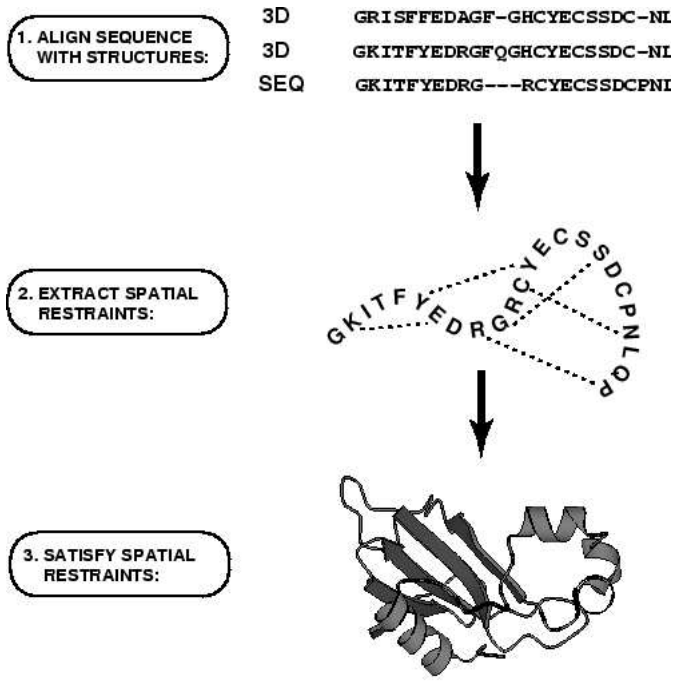


Figure 2.1: The basic methodology followed by MODELLER for comparative modelling by satisfaction of spatial restraints. The target sequence is aligned with the templates, the restraints extracted from the crystal structures and the model is built and optimised (from the MODELLER 6 manual, http://salilab.org/modeller/manual).

parameters allows for the extra optimization of the model through molecular dynamics. MODELLER also allows for the addition of extra restraints such as enforcing secondary structure, incorporating results from cross-linking experiments and NMR results, which can help to improve the model. Further discussion will relate to MODELLER, as it was used to construct the models in this study.

Loop modelling is one of the most difficult areas of homology modelling. Some enzymes differ only by the length of the loops between the secondary structure elements and these loops are more often than not also responsible for the specificity of the enzyme (Marti-Renom *et al.,* 2000). Various methods have been used to try and model loops such as database searches of known loop structures (Chothia *et al.*, 1987; Greer, 1980), molecular dynamics simulations (Bruccoleri *et al.*, 1990; Fiser *et al.*, 2000), genetic algorithms (McGarrah *et al.*, 1993; Ring *et al.*, 1994), biased probability Monte Carlo search (Evans *et al.*, 1995; Thanki *et al.*, 1997) and systematic conformational searches (Brower *et al.*, 1993; Bruccoleri *et al.*, 1993) to mention but a few. MODELLER once again tries to optimize the objective score of the loop to improve the model.

Another issue encountered when modelling sequences with low similarity is that of side-chain conformations. When two sequences have a 30% or less similarity, side-chain packing and interactions generally become unpredictable. Exposed side-chains also present a problem as they may adopt various different conformations without a single one being dominant (Chothia *et al.*, 1987; Rost, 1999).

As with any theoretical method there are certain restrictions during comparative modelling. Errors can be introduced in the protein backbone as a result of insertions (areas in the target without a template match) or deletions in the target sequence. Misalignment is one of the only errors which can be corrected to a certain extent. Sequence diversity will also introduce side-chain packing and interaction errors as well as using wrong templates in cases of distantly related proteins (Marti-Renom *et al.,* 2000).

Homology modelling (or comparative modelling) with MODELLER is divided into four steps: (1) fold assignment and template selection, (2) template-target alignment, (3) model building and (4) model evaluation and refinement. Template selection and fold assignment are some of the most important steps in comparative modelling. Various methods such as BLAST (Altschul *et al.,* 1990) and PSI-BLAST (Altschul *et al.*, 1997) can be used to identify templates from sequence and protein databases such as SCOP (Hubbard *et al.*, 1999) and the Protein Data Bank (PDB). When template-target similarity is low, methods such as threading or 3D template matching methods can be used to identify the fold. Threading methods rely on pairwise comparisons of the template and a protein structure with subsequent scoring of the resulting fit. When fold assignment is difficult due to low sequence similarity, a preliminary model can be built with threading methods. This preliminary model can then be evaluated based on known properties of the protein such as interaction sites, active site residues and conserved residues (Marti-Renom *et al.,* 2000).

To achieve a reliable model an optimal alignment between the templates and target is needed. Alignments are produced from the database hits when searching for templates but these are more often than not sub-optimal. To achieve optimal alignments and hence accurate models, the sequence alignment needs to contain as many related sequences as possible. This will ensure that the globally conserved regions are identified and that this information is incorporated into the template-target alignment. Modern methods such as Hidden Markov Models (HMM) have been used to create models of protein families to help in identifying more remote homologs (Grundy *et al.,* 1997). HMM's have been used to construct the Pfam database, which contains HMM's of all the known protein families (Bateman *et al.*, 2004). To create an HMM, a sequence alignment of all the known sequences are made and the resulting HMM is used to search for new, related proteins. By aligning the target with the alignment used in creating the specfic HMM, the globally conserved regions in the target can be identified. This helps in aligning the template and target in local regions where there is no or low sequence similarity. More modern methods such as 3D T-coffee use information contained in the crystal structure to help align the sequences (Poirot *et al.*, 2004). This ensures that in local, low similarity areas, alignment gaps are not introduced in structurally conserved regions. Another method which can help to improve alignments is secondary structure prediction. Running multiple secondary structure prediction algorithms on the template sequence will help in achieving a consensus of the predicted secondary structure. The predicted secondary structure can help to improve alignments in areas where there is low sequence similarity between target and template.

One or more templates can be used to build a model but in order to avoid small structural errors in the model only closely related structures should be used (if possible).

After the model has been built it needs to be evaluated. The first and most basic step is to check if the model has the correct fold. The wrong fold can usually be avoided by using a closely related template to start with. Once the fold has been confirmed, local areas can be analysed. If the sequence similarity is above 30% the correct folding of local areas and even some chains can be assumed (due to the higher sequence similarity). Once again low similarity (<30%) sequences will have to be checked for correct side-chain orientations. There are various programs which check some basic stereochemistry (PROCHECK, Laskowski *et al.*, 1993; WHATCHECK, Hooft *et al.*, 1996). Lazaridis and Karplus (Lazaridis *et al.*, 1999) used the CHARMM vacuum potential together with a Gaussian distribution model for solvation free energy to show that correctly folded states are always more stable than misfolded states. Their conclusion was that a molecular mechanics energy function coupled to a simple solvation free energy model can perform as well as statistical methods to discriminate between correctly and incorrectly folded proteins. For this study we used PROCHECK.

There are a few limitations to homology models. The resolution of the models cannot go down to atom level (except possibly in high homology cases where sequence similarity is >70%). Another limitation is that homology modelling cannot yet do *ab initio* modelling of sequences longer than 7 residues. The next section will describe the implementation of the discussed methods on the *P. falciparum* PPPK and DHPS sequences.

2.2. Methods

Several steps need to be followed to ensure an accurate homology model of a protein. The first is a correct multiple sequence alignment for all the known protein sequences from the desired family. This ensures that any globally conserved regions are identified. The multiple sequence alignment is then used to achieve an accurate template-target match for the modelling process. The second step is to predict the secondary structures of any regions which do not have a template match. Incorporating the predicted secondary structure can help in increasing the overall usefulness of the model. A third step is to include co-crystallized ligands and co-factors in the model. This helps to restrain the active site residues to the correct conformations. The application of these steps to *P. falciparum* PPPK and DHPS will be discussed in the following sections in more detail.

2.2.1. Homology modelling of *P. falciparum* DHPS

2.2.1.1. Plasmodial DHPS Sequence Retrieval

A Swiss-Prot/TrEMBL (http://www.expasy.org/sprot/, Boeckmann *et al.*, 2003) search for Plasmodial DHPS revealed partial sequences for two members of the *Plasmodium* family (*P. falciparum* and *P. chabaudi*). Both of these entries have been classified as partial sequences (they have been translated from genomic sequences and have not been confirmed) but they seem to contain the complete DHPS sequence. The *P. falciparum* DHPS sequence was used as the target sequence for modelling (TrEMBL accession nr: Q27865) and the *P. chabaudi* sequence (TrEMBL accession nr: Q9BLN5) was added to aid in the multiple sequence alignment. To obtain sequences not yet in Swiss-Prot, a tblastx and blastp search was done against the NCBI genome database (http://www.ncbi.nlm.nih.gov/Blast.cgi). Tblastx is a modified form of BLAST, which takes a DNA query and translates it into all six reading frames and then queries the results against a translated database. From the tblastx results of the DNA sequence of *P. falciparum* DHPS, the sequence of *P. berghei* DHPS (Genbank accession nr: AF344661) was found. From the blastp search the sequence of *P. yoelli yoelli* (Genbank accession nr: EAA21661) was also obtained. Towards the end of the project (2004) Korsinczky *et al.* (2004) published the sequence of the *P. vivax* PPPK-DHPS gene (Genbank accession nr: AY186730). This sequence was not included in the orignal alignment, however, it was used later during the studies of the parasite-specific insertions and their conservation.

2.2.1.2. Multiple Sequence Alignment

An accurate multiple sequence alignment of DHPS sequences is needed in order to deduce the correct conserved regions between all species. Pfam (Bateman *et al.*, 2004) is a large database of protein family alignments and Hidden Markov Models (HMM) derived from the multiple sequence alignments of all the known sequences for protein families. Pfam provides the HMM as well as a choice between the complete multiple sequence alignment of the family or a core multiple sequence alignment (also called a seed alignment) of the family. By using the complete multiple sequence alignment of the DHPS family as a profile, the Plasmodial sequences can be aligned according to conserved positions in all the known sequences. This usually results in very accurate multiple sequence alignments, which incorporate structural conservation (due to the large number of sequences used for the complete multiple sequence alignment).

The *M. tuberculosis* DHPS (MtbDHPS) crystal structure was used as one of the templates for homology modelling as it had the highest identity (30%) to *P. falciparum* DHPS, as well as a vital Mg^{2+} ion in

the active site. *B. anthracis* DHPS (BaDHPS) was also used as it contained a product analogue in the active site. The product analogue crystallized into *B.anthracis* DHPS provided the first concrete clues to the location of *p*-aminobenzoic acid in the DHPS active site. The profile alignment mode in ClustalX (Thompson *et al.*, 1994) was used to align the Plasmodial sequences to the DHPS seed alignment from Pfam (PF00809). The resulting alignment was used to align the templates (*M. tuberculosis*, *B. anthracis*) and the target sequence (*P. falciparum*) for homology modelling according to globally conserved features. To help ensure the accuracy of the alignment the sequences were also submitted to MEME (Bailey *et al.*, 1994) to identify conserved motifs throughout all the known DHPS sequences.

2.2.1.3. Secondary Structure Prediction

Proteins usually consist of a combination of α-helices and β-sheets. The secondary structures of the globally conserved regions in *P. falciparum* DHPS were already known based on the alignment but the inserts had no known template match. The *P. falciparum* DHPS sequence was submitted to various secondary structure prediction servers. These included the DSC secondary structure prediction server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_dsc.html, King *et al.*, 1996), the PredictProtein server (http://cubic.bioc.columbia.edu/predictprotein/, Rost *et al.*, 2003), PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html, McGuffin *et al.*, 2000), 3-D Jury server (http:// bioinfo.pl/Meta/, Ginalski *et al.*, 2003) and the HMMSTR/Rosetta server (http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php, Bystroff *et al.*, 2000). All the results from the various servers were combined to obtain a consensus result for the secondary structure of the inserts.

To obtain a better understanding of the parasite-specific inserts as well as the overall structure of PfDHPS, the program SEG (Wootton *et al.*, 1993) was used for low complexity prediction and ProMate (Neuvirth *et al.*, 2004) was used for protein-protein interaction predictions. In this study the inserts were defined as those areas which did not show any alignment matches in the global Pfam alignment.

2.2.1.4. Homology modelling

MODELLER 6v3 (http://salilab.org, Sali *et al.*, 1993) was used to build the homology models. For a first run the DHPS crystal structures of *M. tuberculosis* (Baca *et al.*, 2003, PDB code: 1EYE), *S. aureus* (Hampele *et al.*, 1997, PDB code: 1AD1) and *E. coli* (Achari *et al.*, 1997, PDB code: 1AJ2) were used as templates (the *B. anthracis* structure was not yet available at this time). The result of the multiple template usage was not optimal as some of the crystal structures contained ligands in the active site and some did not. Not all the crystal structures contained all the loops and thus it was later decided to use both *M. tuberculosis* and *B. anthracis* (Babaoglu *et al.*, 2004, PDB code: 1TX0) as templates. This ensured the least amount of residues without a template match while combining a product analogue and a

metal ion in the active site. The templates and the target were aligned according to the multiple sequence aligment obtained in section 2.2.1.2. Ten models (with and without the longer insertion, to investigate differences in regions lacking templates) were then built with MODELLER (hetero atom set to 'ON', water set to 'OFF'), which included the substrate analogue and Mg^{2+} ion from the crystal structure. This ensured that the restraints imposed on the active site by the crystal structure were transferred to the model during model building. In an effort to improve accuracy the extended loop modelling function of MODELLER was used. The validity of the models was judged by using the MODELLER objective function as well as the overall structure of the model. The model with the longer insert only resulted in a peptide chain extending from the model and thus the model without the longer insertion was used for further study. The model with the highest objective score, which did not contain any knots in the protein backbone (knots are where the backbone of the protein forms a loop and then the backbone pass through the loop), was chosen. The objective score is a function implemented by MODELLER to assess models. The objective score function is based on molecular dynamics principles and the program tries to optimise certain parameters for the model. MODELLER was run on a dual Pentium III (1 GHz) cpu, 1GB RAM, running SuSE 9.1.

2.2.1.5. Mutation Modelling

There are at least five known mutations linked with resistance of sulfadoxine to *P. falciparum* DHPS (S436A/F, A437G, K540E, A581G, A613S/T, Triglia *et al.*, 1997). In order to investigate the effect of the mutations, InsightII (Accelrys) was used to produce the mutation-containing models. The original model was loaded into InsightII and then, using the Biopolymer module, was modified to the desired mutations. The following mutation models were constructed:

- S436A/F
- A437G
- K540E
- A581G
- A613T

These models were subjected to 100 steps of minimization (to ensure that no major energy clashes occurred) in InsightII and further used to investigate the effect of the different mutations on substrate and drug orientation in the active site. InsightII was run on a Silicon Graphics Inc. Octane2 workstation with 512MB RAM.

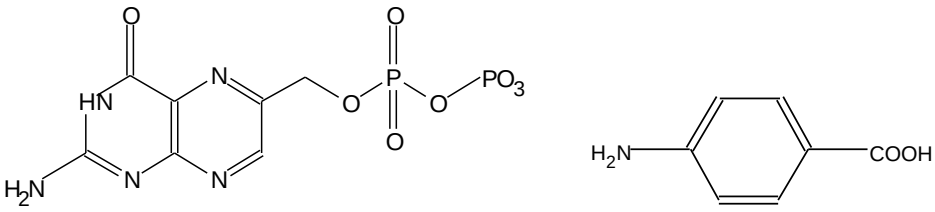


Figure 2.2: The structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (DHP) (left) and *p*-aminobenzoic acid (right). The amino group on *p*-aminobenzoic acid attacks the carbon atom linking the pyrophosphate group and the ring system on DHP (Baca *et al.*, 2000).

2.2.2. Ligand Docking in *P. falciparum* DHPS

The *M. tuberculosis* crystal structure contained a substrate analogue and Mg^{2+} ion, and *B. anthracis* contained a product analogue. To obtain a global picture of ligand-protein interaction, the orientation of the natural substrates and sulfadoxine in the wild-type and the resistant types needed to be compared. This was achieved by building the ligands into the model from the crystal structures with MODELLER, or docking them into the active site model using docking software such as the Ligand Fit module in Cerius2 (Accelrys). The product analogue was modified to the two natural subtrates (using InsightII) and sulfadoxine was docked into the active site using Cerius2 (Accelrys).

2.2.2.1. Substrate Construction

During the modelling process, the Mg^{2+} ion in *M. tuberculosis* DHPS and the product analogue in *B. anthracis* DHPS were incorporated into the model for PfDHPS. The model was loaded into InsightII and the product analogue was modified to the natural substrates, 6-hydroxymethyl-7, 8-dihydropterin pyrophosphate (DHP) and *p*-aminobenzoic acid (Figure 2.2). The substrates were saved as .pdb files for later use in ligand docking.

2.2.2.2. Sulfa-drug Construction and Docking

Various sulfa-drugs are in use against parasites but the most commonly used drug against malaria is sulfadoxine (Figure 2.3). Most of the sulfa-drugs are *p*-aminobenzoic acid analogues with various functional groups attached. The following sulfa-drugs were constructed with the Builder module of InsightII and saved as .mol2 files for later use: sulfathiazole, dapsone, sulfachloropyridazine, sulfaquinoxaline, sulfadimethoxine, sulfapyridazine, sulfamoxole, sulfamethoxypyridazine, sulfamethoxazole, sulfamerazine, sulfadiazine, sulfisoxazole, PAS, sulfanilamide and sulfacetamide (Kasekarn *et al.*, 2004). The constructed sulfa-drugs were minimized in InsightII until the change in energy was less than 0.01 kcal/mol using the CVFF forcefield (Accelrys).

Only sulfa-drug resistance with respect to sulfadoxine was investigated and not the detailed docking of the other sulfa-drugs. For the docking of the sulfa-drugs, the LigandFit module of Cerius2 (Accelrys) was

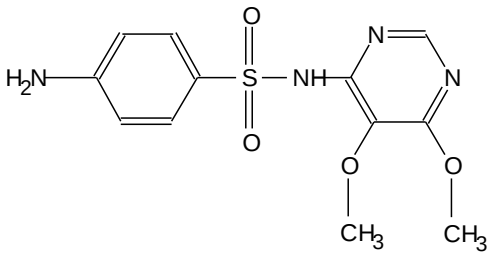


Figure 2.3: The structure of sulfadoxine, a *p*-aminobenzoic acid analogue (Kasekarn *et al.*, 2004).

used. The model was loaded with the substrates in place (as constructed in InsightII). The Cavity Finding module was used to identify all the cavities and the active site cavity was selected. The following programs were also used in attempts to dock sulfadoxine: the LUDI module of InsightII, Autodock (Morris *et al.*, 1998) as well as GOLD (Jones *et al.*, 1997). Cerius2 LigandFit produced 20 fits with a range of conformations. The orientation in which the *p*-aminobenzoic acid part of sulfadoxine showed the closest resemblance to the crystal struture orientation was chosen. Only Cerius2 succeeded in docking sulfadoxine in an acceptable conformation and orientation. An acceptable conformation was a conformation in which sulfadoxine would interact with DHP in a way similar to the PfDHPS catalytic mechanism.

The following parameters were used and applied during the docking of sulfadoxine into the DHPS model with Cerius2:

- Hydrogens were automatically checked to ensure that all residues had the correct number of hydrogens.
- The automatically identified binding pocket was enlarged to ensure an adequate sampling of the conformational space in the active site.
- The default grid-resolution size was kept at 0.5Å to give an accurate representation of the active site.
- A pore opening of 8Å was specified to ensure that an adequate region of conformational space was sampled.
- The option to penalise any orientation of sulfadoxine if some part of the molecule was beyond the defined active site, was activated.
- Cerius2 was allowed to make the ligand flexible.

The sulfadoxine-PfDHPS complex was minimized according to the same criteria used in section 2.2.2.2 for minimizing the ligand-PfDHPS complex. The interaction between the DHPS model and the substrates was then visualized using LIGPLOT (Wallace *et al.*, 1995). The interaction between DHPS and sulfadoxine was compared to the interactions published for *M. tuberculosis, B. anthracis, S. aureus* and

E. coli DHPS (Baca *et al.*, 2000). Cerius2 was run on a Silicon Graphics Inc. Octane2 workstation with 512MB RAM.

2.2.3. Homology modelling of *P. falciparum* PPPK

2.2.3.1. Multiple Sequence Alignment

To ensure a reliable alignment of PfPPPK with the template, the PPPK (Pfam reference: PF01288) profile from Pfam was used. This ensured a reliable alignment of the globally conserved regions. The complete *P. falciparum* PPPK sequence from Genbank (Genbank accession nr: U07706) was used as well as partial sequences from *P. vivax* (Genbank accession nr: AY186730), *P. berghei* (Genbank accession nr: AF344661) and *P. chabaudi* (Genbank accession nr: AJ302077), which were obtained through a tblastx search with the *P. falciparum* PPPK DNA sequence. Blastp provided the PPPK sequence of *P. yoelli yoelli* (Genbank accession nr: EAA21661). The globally conserved regions were used to align the sequences for which a crystal structure was available. PPPK of *E. coli* (PDB ref: 1DY3, 1CBK, 1EQM, 1EQO, 1EX8, 1F9H, 1F9Y, 1G4C, 1HKA, 1HQ2, 1IM6 and 1KBR) and *H. influenzae* (PDB ref: 1CBK) had available crystal structures. PfPPPK showed 21% identity and 40% similarity to EcPPPK.

2.2.3.2. Homology modelling

A reliable structural model needs to include as many restraints as possible when building the model. All the available crystal structures of PPPK were investigated. There are 12 PPPK crystal structures available (11 from *E. coli* (Ec; Stammers *et al.*, 1999; Blaszczyk *et al.*, 2000) and one from *H. influenza* (Hi; Henning *et al.*, 1999)). The *E. coli* crystal structure was used as it contained the most complete set of substrates (ATP and a pterin substrate analogue, 7,8-dihydro-6-hydroxymethyl-7-methyl-7-pterin) and co-factors (two Mg^{2+} ions). The other crystal structures (PDB ref: 1CBK, 1EQM, 1EQO, 1EX8, 1F9H, 1F9Y, 1G4C, 1HKA, 1HQ2, 1IM6 and 1KBR) contained various forms of substrate analogues and either ADP or AMP. From the multiple sequence alignment it was evident that *P. falciparum* PPPK contained two large inserts (each about 90 residues in length when compared to non-Plasmodial species) and the decision was made to exclude them from the model building process due to their length and the lack of a template match. MODELLER 6v3 was used to build the models. The substrate analogue (7,8-dihydro-6-hydroxymethyl-7-methyl-7-pterin) and ATP, as well as the two Mg^{2+} ions from the crystal structure, were included during the model building. The models were built following the same criteria used for the modelling of *P. falciparum* DHPS in section 2.2.1.4. The substrate analogue

was converted to the natural pterin substrate using InsightII. To obtain a better characterization of the protein, low-complexity region prediction was done using the program SEG (Wootton *et al.*, 1993).

2.3. Results

The following results were obtained from the homology modelling of PfDHPS and PfPPPK.

2.3.1. Homology Modelling of DHPS

The template-target alignment was derived from the Pfam DHPS multiple sequence alignment and the additional four Plasmodial species were added later. Figure 2.4 shows *P. falciparum* DHPS aligned with the four other Plasmodial DHPS sequences as well as the *M. tuberculosis* and *B. anthracis* DHPS sequences. Notice the two inserts when comparing the five Plasmodial species to *M. tuberculosis* and *B. anthracis*. The length of the inserts also differ between the five Plasmodial species but there is a high level of conservation between the Plasmodial species from residues 740-753 in the second insert. The length of insert 1 varies between the Plasmodial species and it does not seem to have any conserved regions. The alignment shows that there is a highly conserved, *Plasmodium*-specific region at the end of insert 2 (residues 747-759 in the *P. falciparum* sequence), which was predicted to be an α-helix. Figure 2.5 shows the conserved motifs in known DHPS sequences predicted by MEME. The alignment suggested that there are two inserts when comparing the DHPS of the Plasmodial species to the rest of the known DHPS sequences. The inserts in the *P. vivax* and *P. falciparum* sequence are shown in Figure 2.5. Motif 13 seems to be missing in *P. falciparum* but this is due to an insert in the junction region in the *P. falciparum* sequence.

Low complexity prediction using SEG, showed no low complexity regions in the *P. falciparum* DHPS sequence. Modeller produced 10 models and the best was chosen according to the criteria listed in section 2.2.1.4. The DHP and the Mg^{2+} ion were built directly into the model from the crystal structure. Figure 2.6 shows interactions in the DHPS model active site. The primary active site residues show a very high degree of conservation when compared to the available crystal structures (Table 2.1). An advantage of using homology modelling is that the active site residues of the model are restrained by the orientation of the template residues, thus providing a very accurate representation of the orientation of the active site residues. DHPS occurs as a TIM barrel and the model also assumed this fold. Figure 2.7 shows the comparison between the *P. falciparum* model, *M. tuberculosis* and *B. anthracis* crystal structures. The main difference between the two models can be seen in the loop regions where *P. falciparum* has loops which are a few residues longer. Figure 2.8 shows the orientation of the substrate, DHP, and the Mg^{2+} ion in the model. Note the proximity of the amino group of *p*-aminobenzoic acid to the carbon

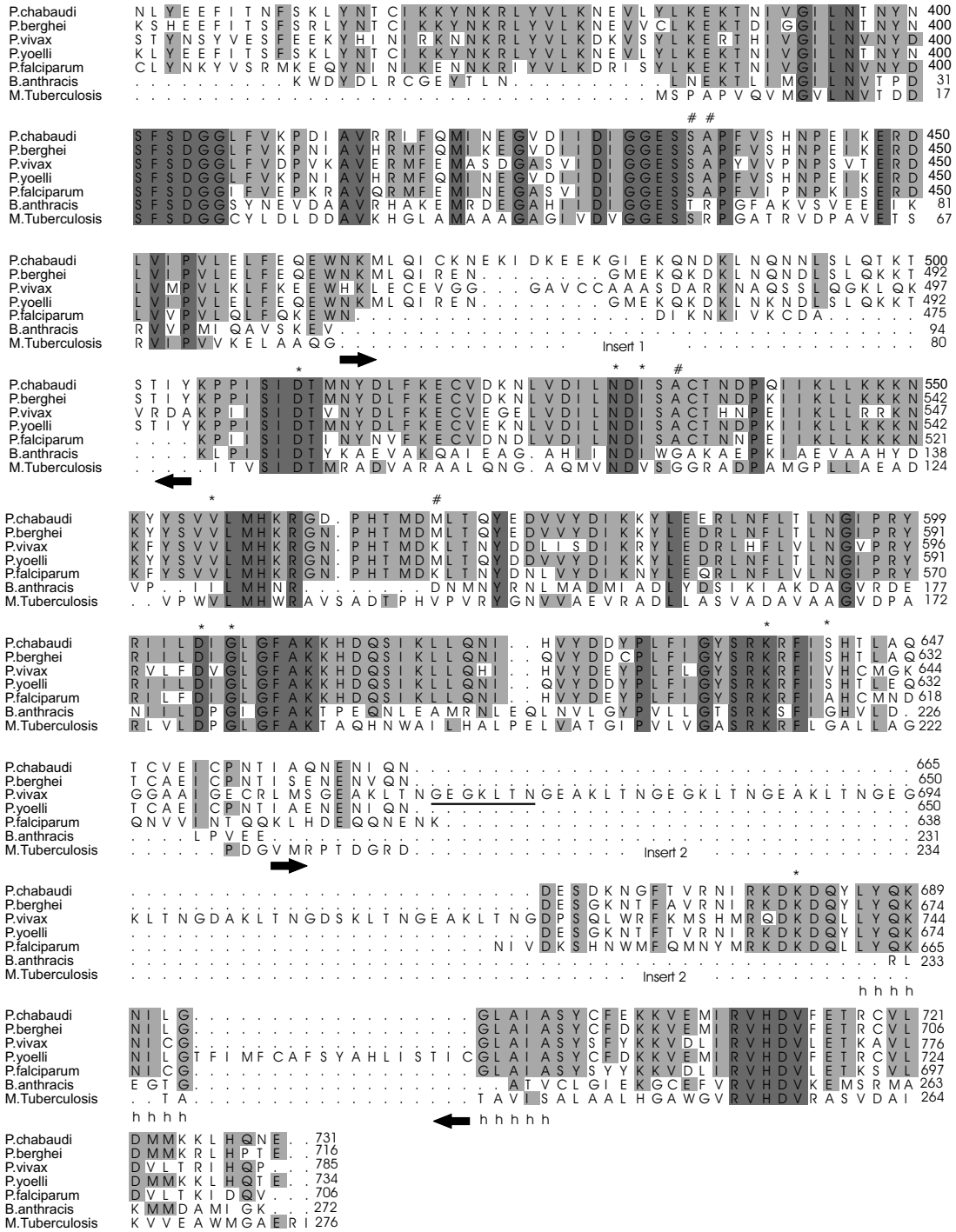


Figure 2.4: A multiple sequence alignment between *B. anthracis*, *M. tuberculosis* and five Plasmodial DHPS sequences (*P. chabaudi*, *P. berghei*, *P. yoelli yoelli*, *P. falciparum* and *P. vivax*). As DHPS forms part of the PPPK-DHPS gene, the DHPS sequence numbers for the *Plasmodium* species start at 350, thus maintaining the original numbering for the bifunctional protein. The arrows indicate the start and end of the insertions, the underlined region is the repeat present in *P .vivax*, 'h' indicates the predicted α-helix in insert 2, '#' indicates the known resistance-causing mutations and '*' indicates residues involved in direct substrate contact (Baca *et al.*, 2000).

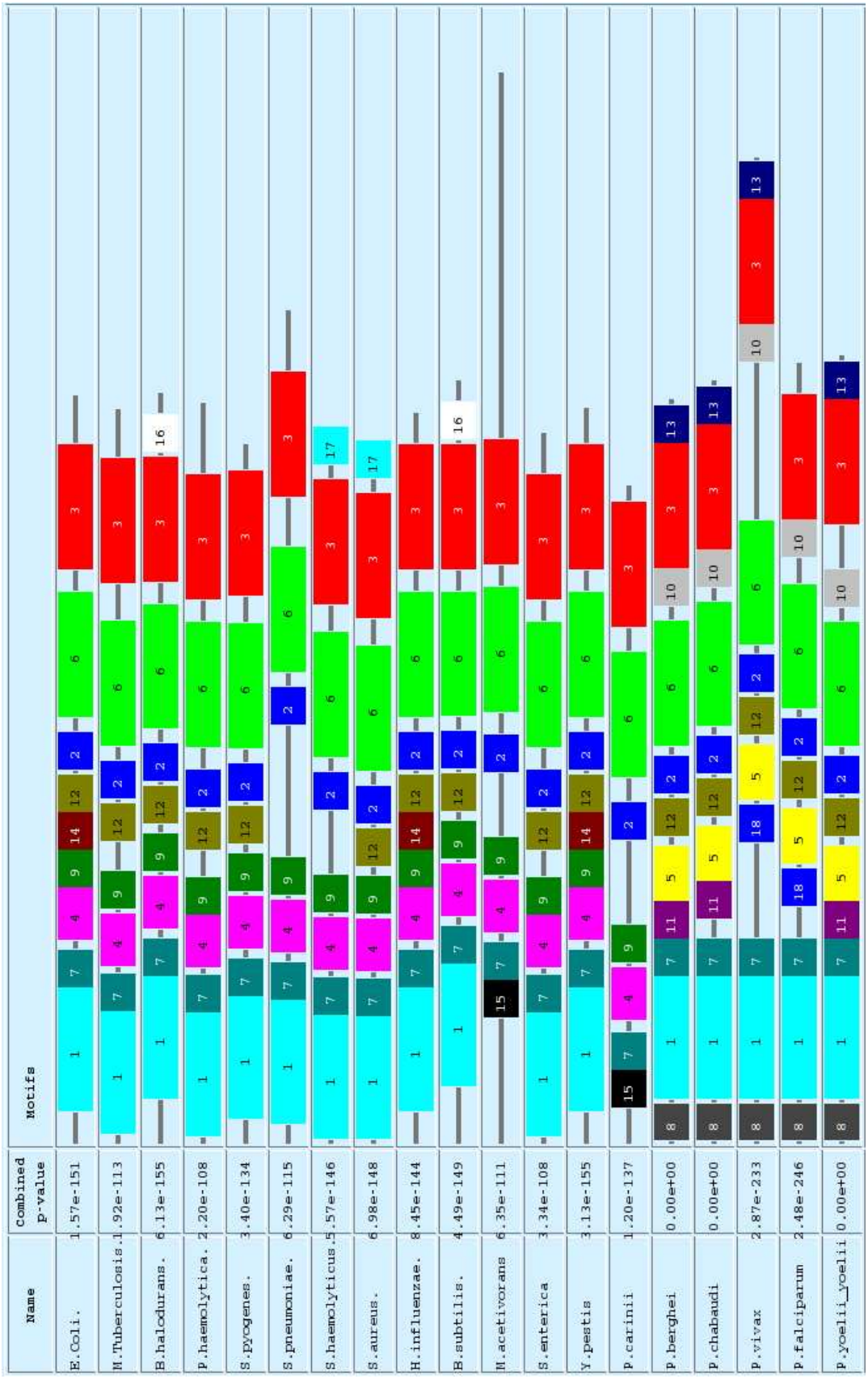


Figure 2.5: The conserved motifs in DHPS identified by MEME. Note the extra motifs 5, 8 and 10 identified by MEME. These motifs are unique to the *Plasmodium* species. The *Plasmodium* species also lack motif 4 and 9. Note the extra motif 11 in the sequences of *P. berghei*, *P. chabaudi* and *P. yoelli yoelli*. Motif 11 supports the hypothesis that these three species are closely related (Kedzierski *et al.*, 2002). *P. falciparum* also lacks motif 13 as it seems to contain an insert in the junction region when comparing it to the other *Plasmodium* species.

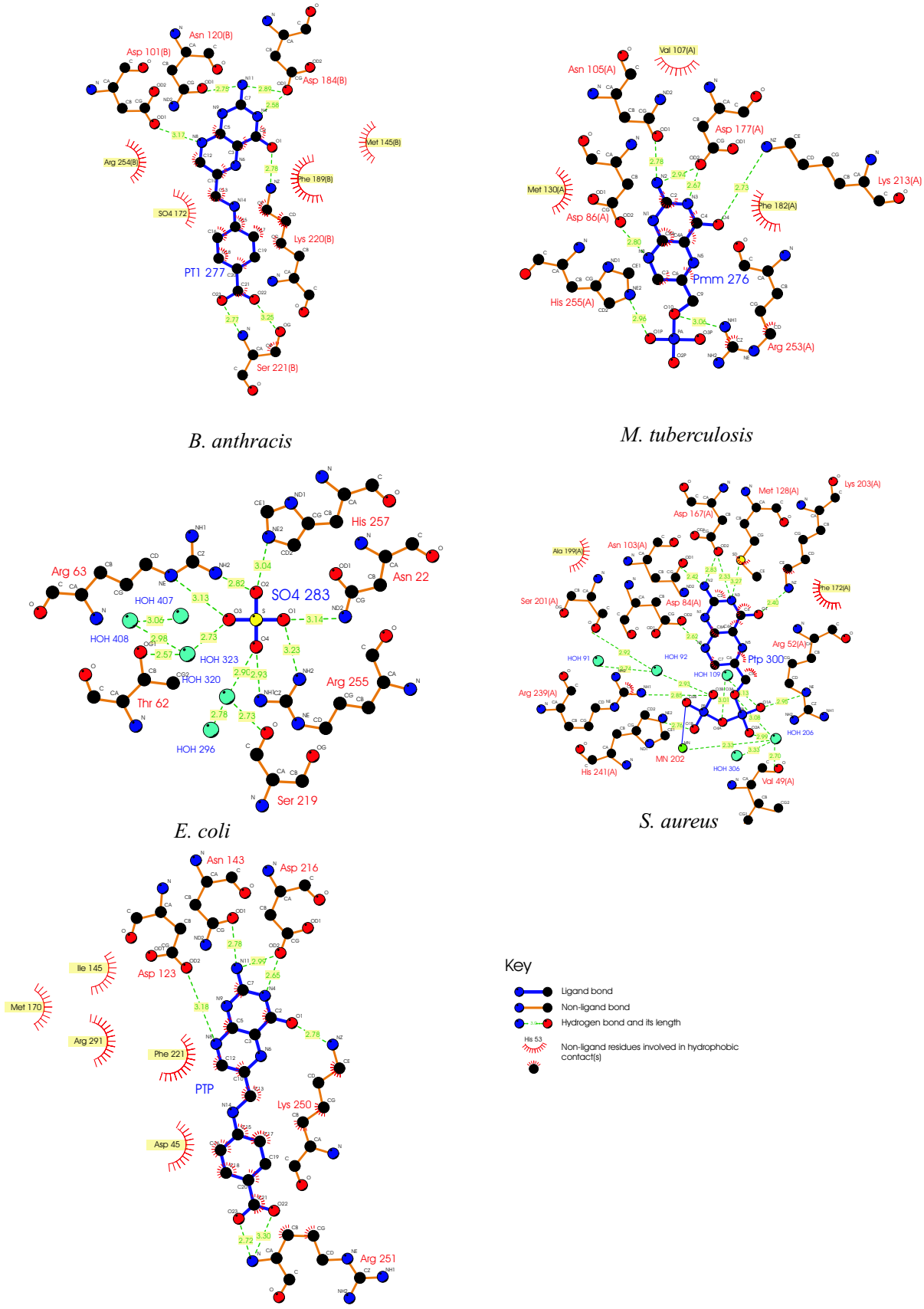


Figure 2.6: Contacts in the active site of the *P. falciparum* DHPS model in comparison to the known DHPS crystal structures. PT1, PTP and PMM indicate the positions of the substrate (or its analogues) in the known crystal structures. The numbering are those in the crystal structures (or in the case of *P. falciparum*, the model), PT1: pteroic acid, Pmm: pterin-6-yl-methyl-monophosphate, SO_4: sulfate ion, Ptp: pterin pyrophosphate. All figures produced by LIGPLOT (Wallace *et al.*, 1995).

connecting the pyrophosphate group to the ring system of DHP. During catalysis the amino group attacks this carbon and condenses with DHP. This causes the pyrophosphate group to be released from DHP, resulting in the formation of dihydropteroate. The model contained 9 α-helices and 8 β-sheets similar to the 9 α-helices and 8 β-sheets in the *M. tuberculosis* and *B. anthracis* crystal structures.

The PROCHECK quality check of the PfDHPS model is shown in Figure 2.9. The results for the Ramachandran plot and the estimated accessibility plots are shown. The model showed three residues in the disallowed regions. PROCHECK were also ran on both templates and compared to the model.

Table 2.1: Comparison of the active site residues between *M. tuberculosis*, *S. aureus*, *E. coli*, *B. anthracis* and the *P. falciparum* DHPS model. These residues are highly conserved in the alignment (Figure 2.4) (Baca *et al.*, 2000; Hampele *et al.*, 1997; Achari *et al.*, 1997; Babaoglu *et al.*, 2004).

M. tuberculosis	***S. aureus***	***E. coli***	***B. anthracis***	***P. falciparum****	**Contacts**
Asp 21	Asp 19	Asp 30	Asp 35	Asp 45/404	Interacts with the α-phosphate group on PTP
Asp 86	Asp 84	Asp 96	Asp 101	Asp 123/482	Has hydrogen bonds with the amino group on PTP
Asn 105	Asn 103	Asn 115	Asn 120	Asn 143/503	Has hydrogen bonds with the N-atoms in PTP
Met 130	Met 128	Met 129	Met 145	Met 170/529	Shows hydrophobic interactions with PTP
Asp 177	Asp 167	Asp 185	Asp 184	Asp 216/575	Hydrogen bonds with the amino group on PTP
Phe 182	Phe 172	Phe 188	Phe 189	Phe 221/580	Undergoes hydrophobic interaction with PTP
Ser 211	Ser 201	Ser 219	Ser 221	Ser 248/607	Interacts with *p*-aminobenzoic acid
Lys 213	Lys 203	Lys 221	Lys 220	Lys 250/609	Interacts with the α-phosphate group through a water molecule, as well providing hydrogen bonds to the 4-oxo group on PTP
Arg 253	Arg 239	Arg 255	Arg 254	Arg 291/686	Interacts with the α-phosphate group on PTP
His 255	His 241	His 257	His 256	His 293/688	Interacts with the α-phosphate group on PTP

* The first number is the residue number as it appears in the model. The second number indicates the residue number in the full bifunctional protein.

2.3.2. Ligand Docking

Sulfadoxine was docked into the active site of PfDHPS with the LigandFit module of Cerius2. The *B. anthracis* crystal structure revealed a probable position of *p*-aminobenzoic acid. As sulfadoxine is a *p*-aminobenzoic acid analogue, the *p*-aminobenzoic acid position helped in eliminating some of the possible orientations suggested by Cerius2. Figure 2.10 shows the orientation of sulfadoxine in the active

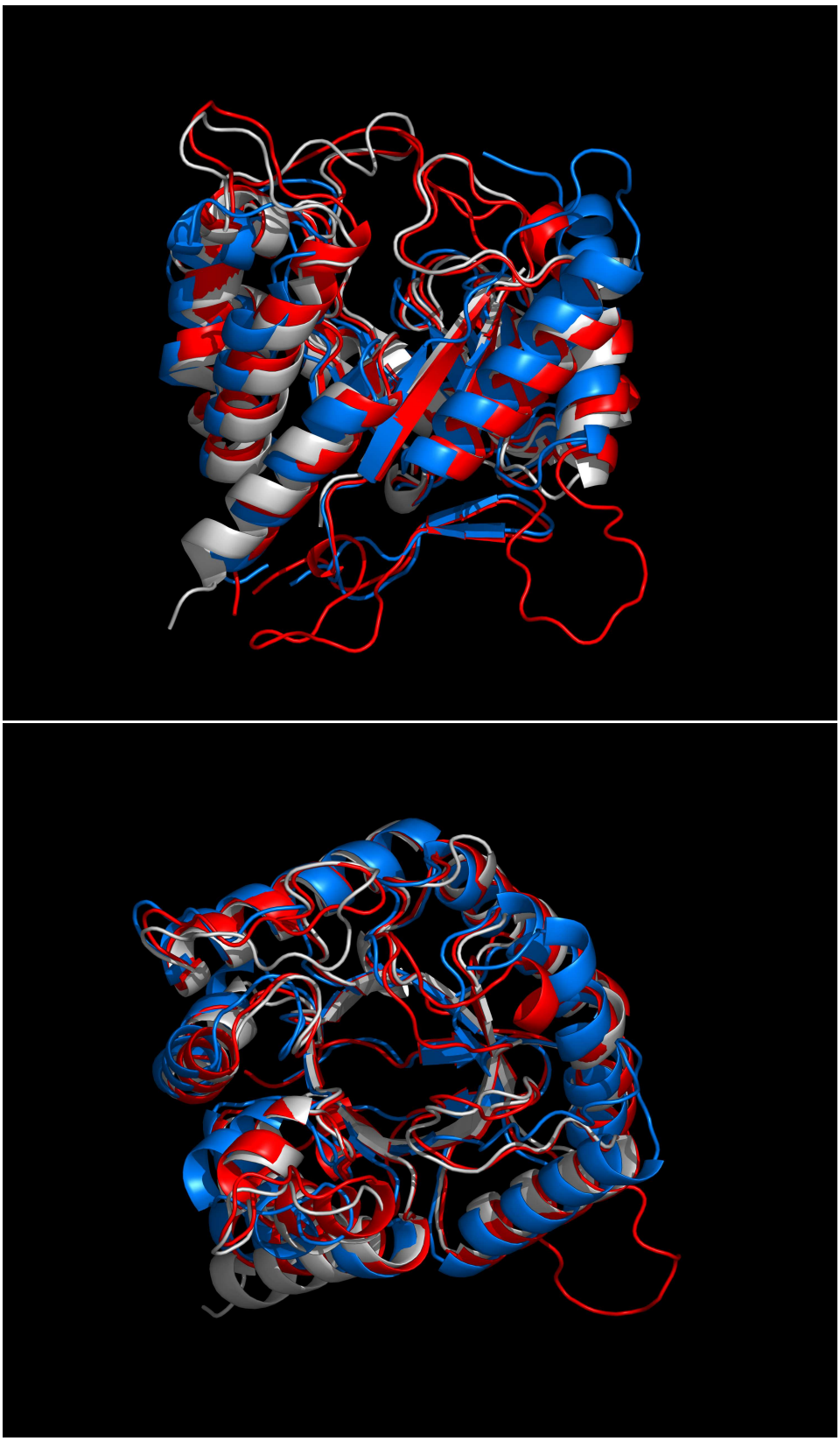


Figure 2.7: A comparison between the backbone structures of *M. tuberculosis* (white), *B. anthracis* (blue) and the *P. falciparum* (red) model. Top: side view of the DHPS model. Bottom: view from the top of the DHPS model. The red loop portruding from the bottom set of models is the shorter insert in the *P. falciparum* sequence (residues 463 to 475 in the bifunctional protein, residues 105 to 117 in the model). Generated with PYMOL (DeLano 2002, http://www.pymol.org).

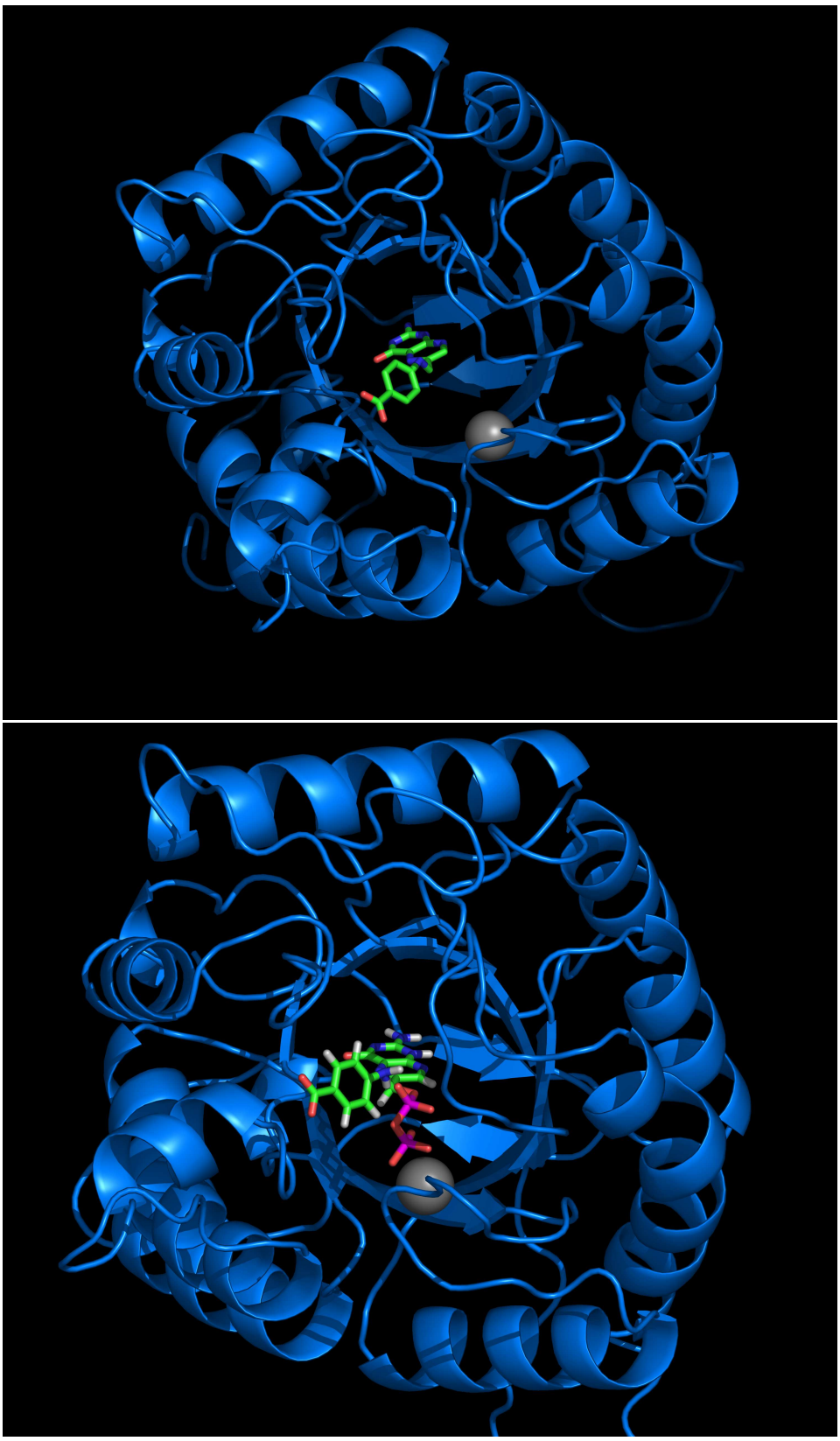
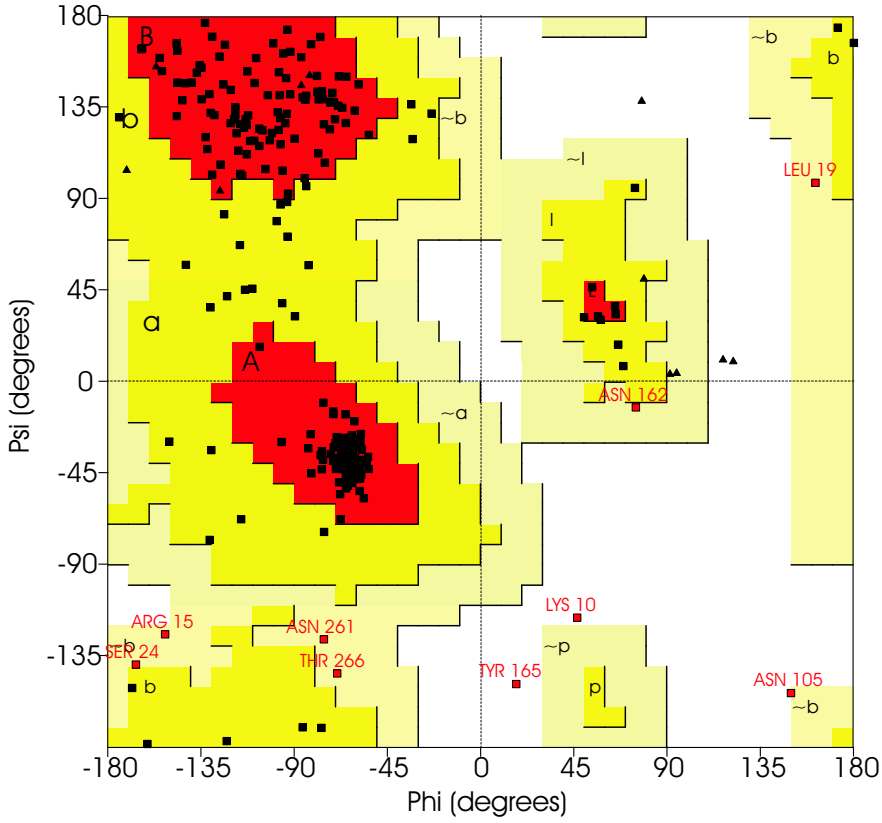


Figure 2.8: The *P. falciparum* DHPS model showing the orientation of the subtsrates and the Mg^{2+} ions. Top: The location of the product analogue and the Mg^{2+} ion (grey) after model construction. Bottom: The DHP substrate and Mg^{2+} ion were built in from the crystal structure during model building, and the *p*-aminobenzoic acid constructed by breaking the bond between the pterin group and the benzene moiety. Generated with PYMOL (DeLano 2002, http://www.pymol.org).



Plot statistics

Residues in most favoured regions [A,B,L]	240	83.9%
Residues in additional allowed regions [a,b,l,p]	37	12.9%
Residues in generously allowed regions [~a,~b,~l,~p]	6	2.1%
Residues in disallowed regions	3	1.0%
Number of non-glycine and non-proline residues	286	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	13	
Number of proline residues	10	
Total number of residues	311	

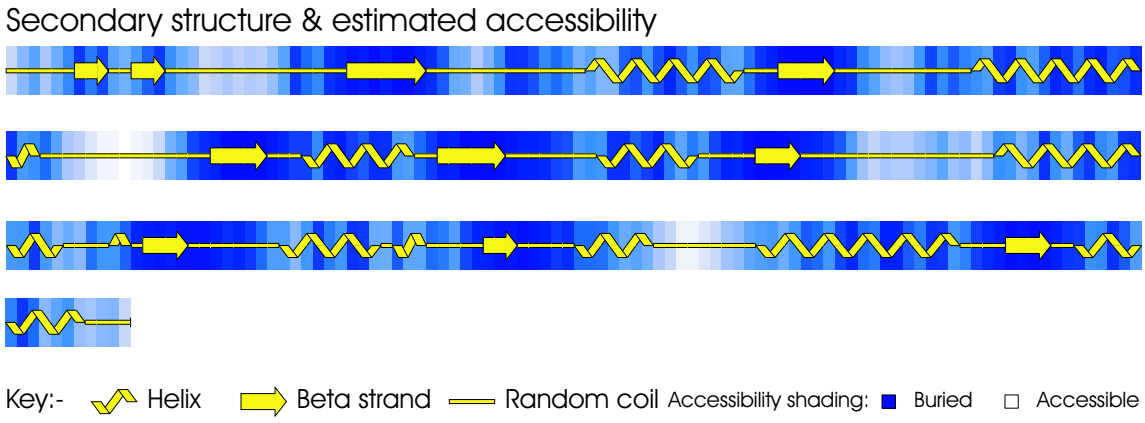


Figure 2.9: Results of the PROCHECK quality check of the PfDHPS model (only selected results are shown). Top: image represents the Ramachandran plot. The three residues in the disallowed regions are a result of the modelling process. During subsequent minimization these residue parameters are adjusted to normal. The PROCHECK results of the temlates showed no residues in disallowed regions. Bottom: An estimated solvent acessibility plot. The plot is an estimate of the solvent exposure of each residue. Each vertical color in the secondary structure plot represents a separate residue (100 residues per plot).

site of PfDHPS. The model indicated that sulfadoxine fills up the remainder of active site cavity (after DHP has docked) and thus blocks the entrance of *p*-aminobenzoic acid into the active site. Sulfadoxine also shows strong interactions with the pyrophosphate group on DHP, which helps to keep it in the active site. The methyl groups on sulfadoxine show hydrophobic interactions with the side chain of Lys 540 in PfDHPS. When Lys 540 is mutated to Glu there is a charge reversal, as well as a shortening of the hydrophobic side chain (Figure 2.10). This may cause a loss of the hydrophobic interaction and thus the methyl groups are forced to move, resulting in a less tight interaction between PfDHPS and sulfadoxine. The A581G simulation showed that the methoxy groups move away from residue 581 due to a decrease in hydrophobic interactions. Ser 436 shows interaction with sulfadoxine and when the S436F/A mutation occurs there is a loss of this interaction. The S436F mutation causes steric hindrance as the Phe mutation takes up the space needed for sulfadoxine to bind. The S436A mutation will cause a loss of interaction as the interaction between Ser 436 and sulfadoxine is through the hydroxyl group of Ser 436.

2.3.3. Homology modeling of PPPK

The template-target alignment was obtained from the multiple sequence alignment and used for modelling. The alignment showed two large insertions (average of 90 residues each) and these inserts were excluded from the model building alignment due to their length as well as the lack of a template match. There is some conservation between the Plasmodial species but the *P. vivax* sequence seems to have diverged the most (Figure 2.11).

Unlike the Plasmodial DHPS sequences, the length of the inserts in the Plasmodial PPPK sequences seems to be more conserved between the Plasmodial species (except for *P. falciparum*). Insert 1 shows a high degree of conservation between the five Plasmodial sequences but insert 2 only shows some conservation between *P. falciparum*, *P. chabaudi* and *P. berghei* and only at the beginning of insert 2. Figure 2.13 shows the results of conserved motif detection in PPPK by MEME. The *Plasmodium* family seems to have undergone huge changes in the PPPK protein as seen by the absence of globally conserved regions. Once again there is evidence for the close phylogenetic relationship between *P. chabaudi, P. berghei* and *P. yoelli yoelli* as seen with motif 15 and 16 only occurring in these three species. *P. yoelli yoelli* seems to contain an extra motif 19 which is similar to that of *C. pneumoniae*. *P. falciparum* also contains an extra motif 17 and 18 when compared to the other members of the *Plasmodium* family.

MODELLER produced 10 models and the models were evaluated based on the criteria in section 2.2.1.4. The substrate analogue 87Y (7,8-dihydro-6-hydroxymethyl-7-methyl-7-pterin) and the Mg^{2+} ions were built in from the template crystal structure and modified to the natural substrate using InsightII (Figure 2.15). The active site of the PfPPPK model compared well with the other known crystal structure active sites (Figure 2.14), with differing residues indicated in Table 2.2. The quality of the model and the templates were checked by PROCHECK and the results for the model is shown in

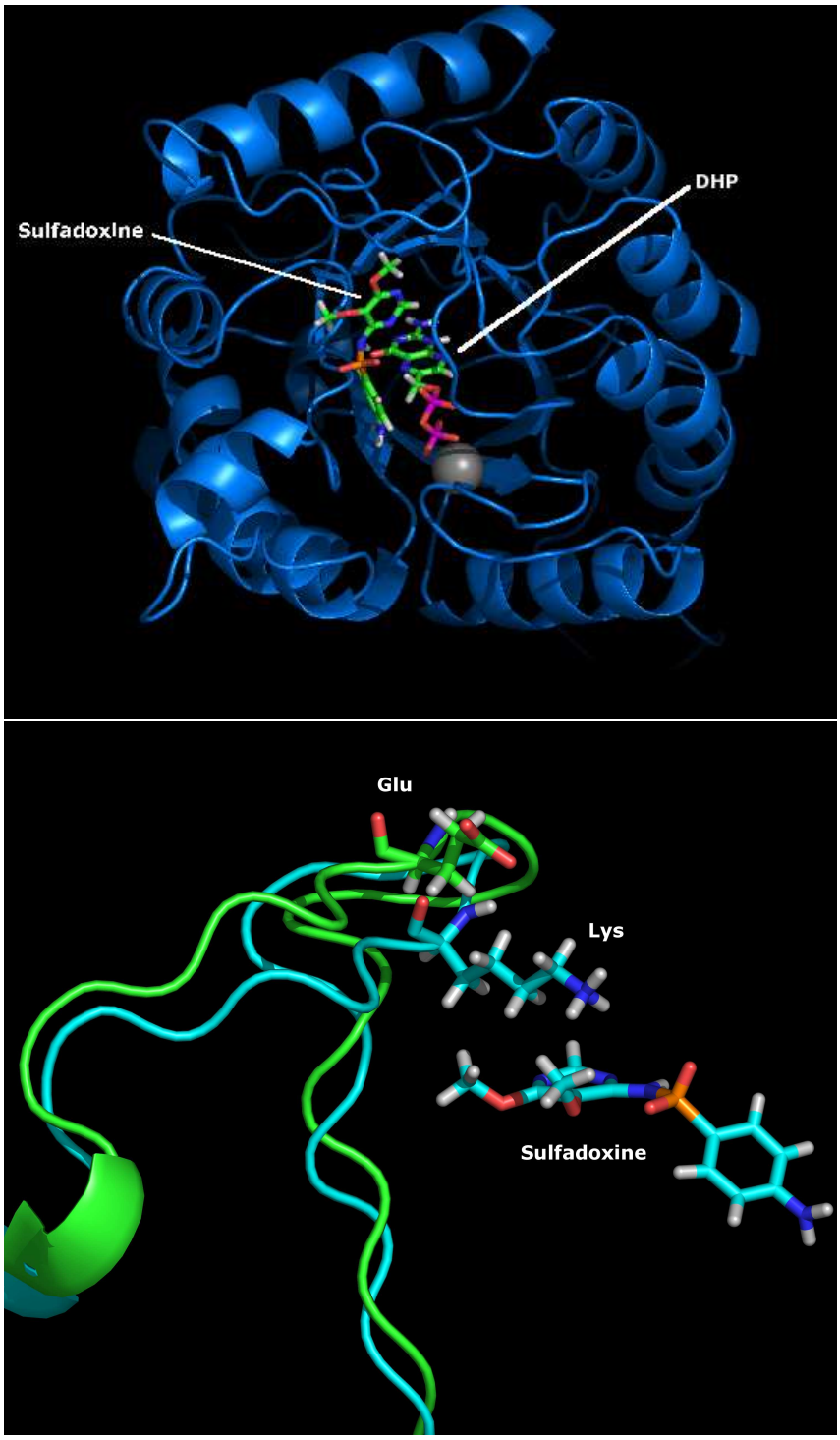


Figure 2.10: Top: The interaction between PfDHPS, DHP and sulfadoxine. The grey sphere is the Mg^{2+} ion, which anchors DHP through the pyrophosphate groups. The *p*-aminobenzoic acid analogue part of sulfadoxine assumes a marginally different orientation than *p*-amino benzoic acid in Figure 2.8 and shows interactions with the pyrophosphate group on DHP. Bottom: The effect of the K540E mutation on sulfadoxine binding. In the wild-type (light blue) the methoxy groups on sulfadoxine shows interactions with the sidechain of Lys 540. When the K540E (green) mutation occurs, the methoxy groups rotate to point away (not shown) and the distance between residue 540 (Glu) and sulfadoxine doubles. This may be due to the increased interaction of Glu 540 with the solvent, which may also explain why the loop opens up more than in the wild-type. Generated with PYMOL (DeLano 2002, http://www.pymol.org).

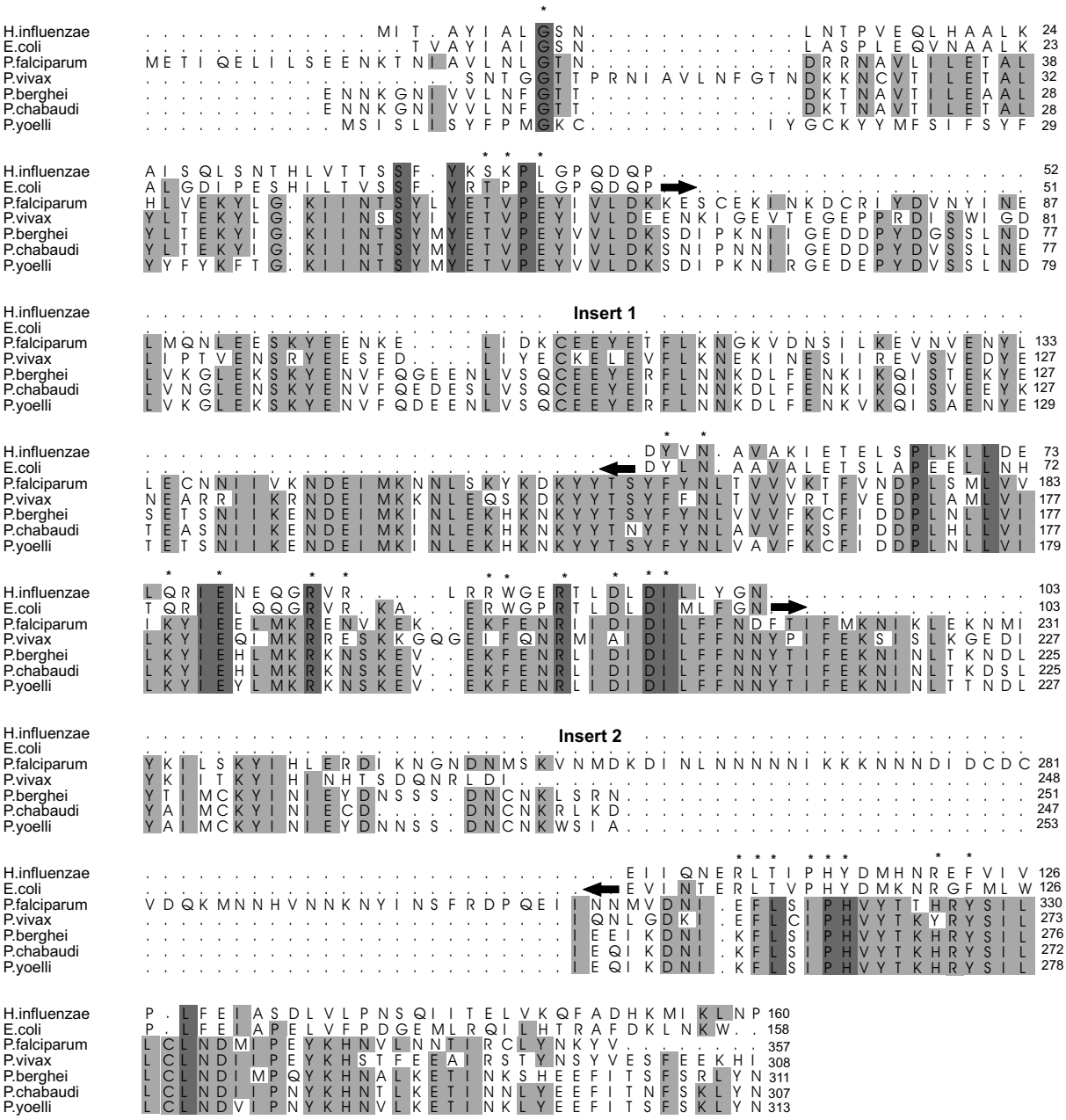
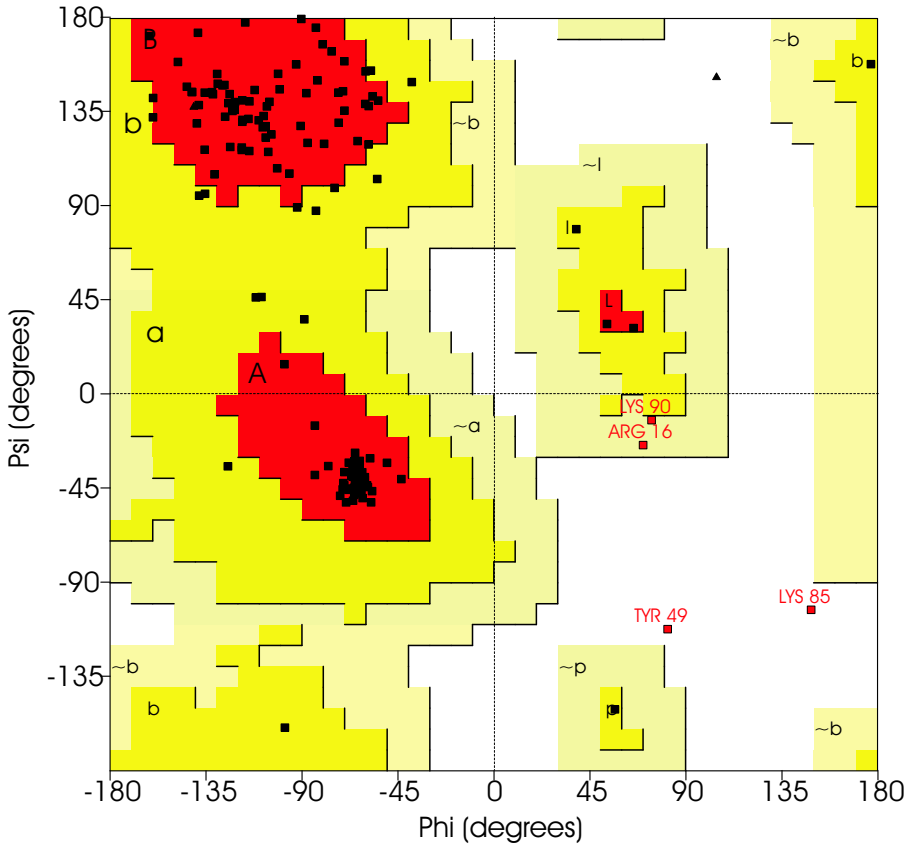


Figure 2.11: The multiple sequence alignment for the complete *E. coli*, *H. influenzae* PPPK and five Plasmodial PPPK sequences. Notice the two large inserts in the *Plasmodium* family when compared to *E. coli* and *H. influenzae*. Insert 1 shows some conservation between the Plasmodial species but insert 2 only has a short conserved motif at the beginning. There are also a few short insertions as seen in insert 1 for *P. berghei*, *P.chabaudi* and *P. yoelli yoelli*. The large inserts (insert 1 and 2) were left out of the modelling process due to their excessive length and the lack of a template match. '*' indicates PPPK contacts with ATP and the substrate analogue. Note the additional *P. falciparum*-specific insert in insert 2. This specific insert is a low-complexity region. PfPPPK numbering stops at 357 and PfDHPS numbering starts at 359, this includes a short linker region between PPPK and DHPS.



Plot statistics

Residues in most favoured regions [A,B,L]	129	85.4%
Residues in additional allowed regions [a,b,l,p]	18	11.9%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.3%
Residues in disallowed regions	2	1.3%
Number of non-glycine and non-proline residues	151	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	2	
Number of proline residues	4	
Total number of residues	160	

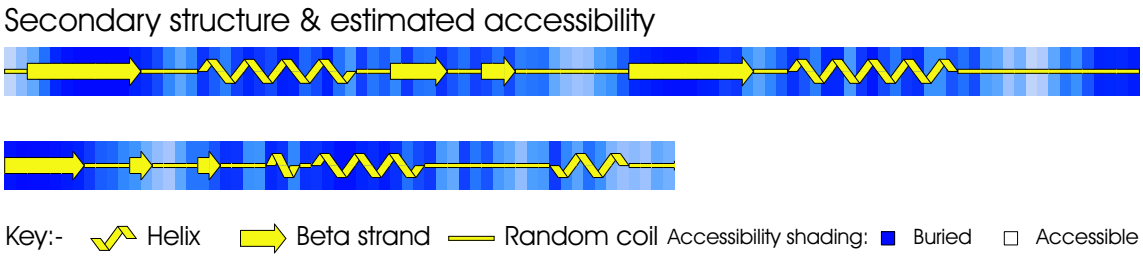


Figure 2.12: Results of the PROCHECK quality check of the PfPPPK model (only selected results are shown). Top: Ramachandran plot of PfPPPK. The two residues in the disallowed regions are due to the modelling process. The residue parameters are shifted to normal during subsequent minimizations. Bottom: An estimated solvent acessibility plot. The plot estimates the solvent exposure of each residue. Each vertical color in the secondary structure plot represents a separate residue (100 residues per plot).

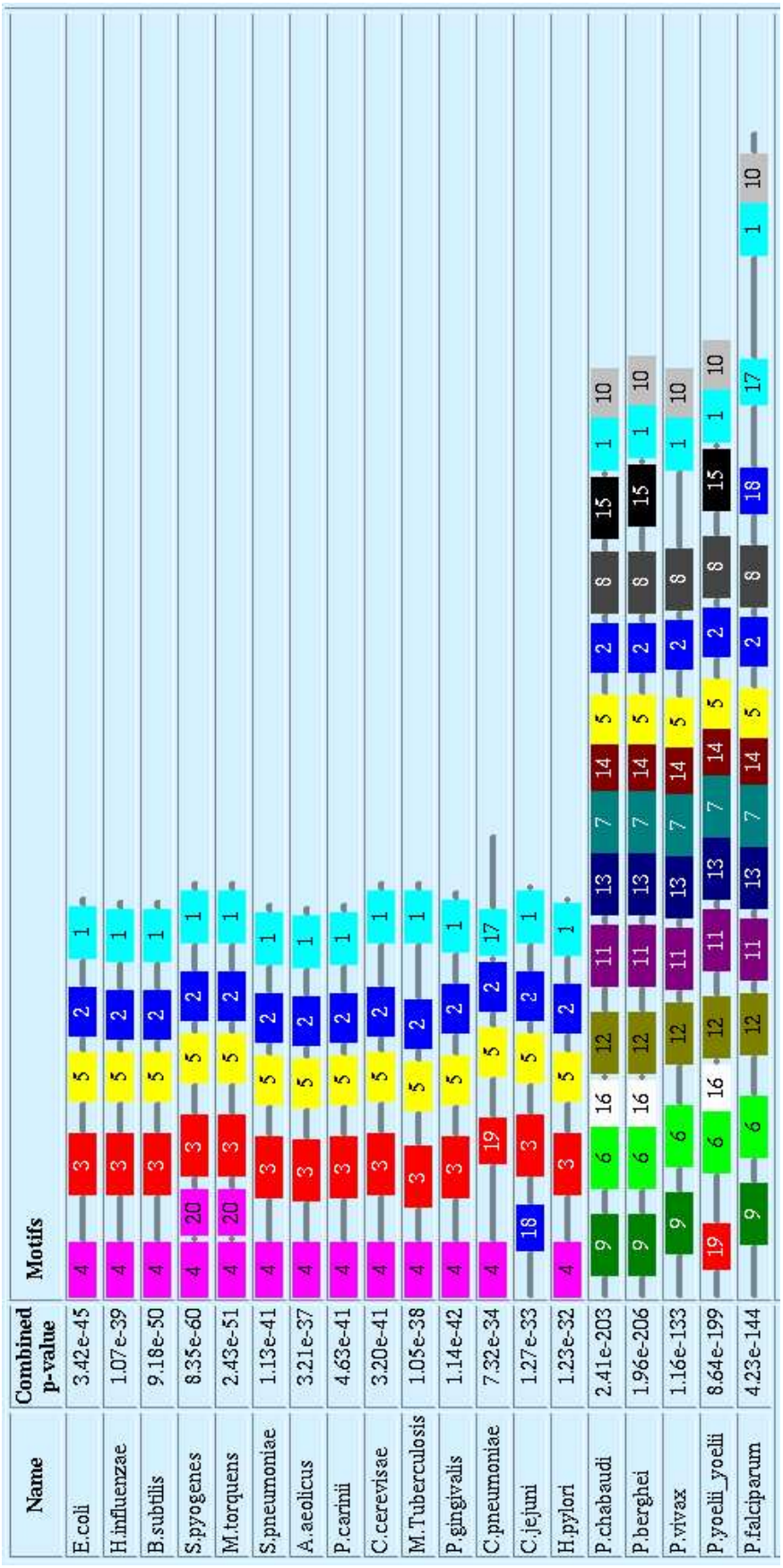
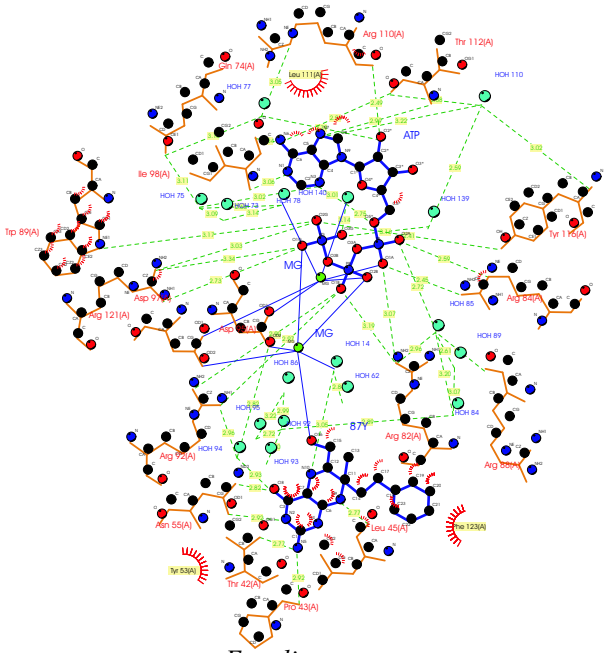


Figure 2.13: The conserved motifs in PPPK as detected by MEME. Note the large inserts in the *Plasmodium* species when compared to the other species.

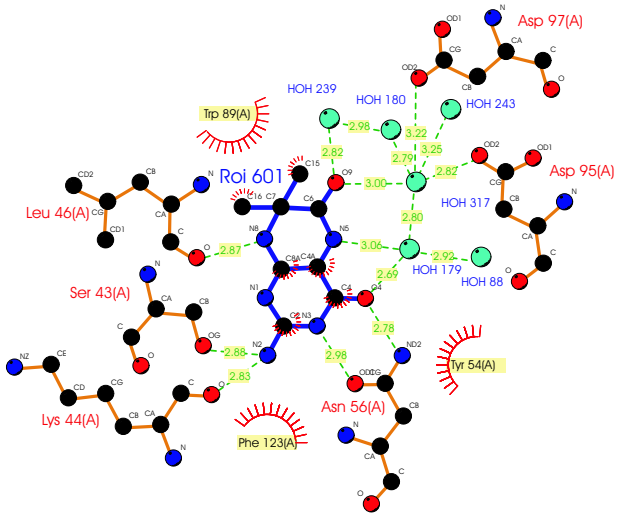
Figure 2.12. During minimization the Mg^{2+} ions anchor the phosphate groups on ATP and prevents it from moving too far away. ATP also remained close enough to the pyrophosphate- accepting group for the reaction to occur. From the crystal structure of *E. coli* it can also be seen that the Mg^{2+} ions play a role in anchoring the phosphate groups of ATP as well as mediating interactions between the protein and the substrate (Figure 2.14).

Table 2.2 shows the interactions between PPPK, substrate and ATP. Blaszczyk *et al.* (2000) remarked that the differences seen in substrate-interacting residues may be responsible for the different isozymes of PPPK isolated from biological sources. These authors also proposed that the following residues are involved in loop stabilization in *E.coli* PPPK: Asn 10, Gln 50, Arg 82, Arg 84, Arg 88, Trp 89, Arg 92 and Tyr 116 (Pf: Asn 25, Asp 65, Asn 100, Glu 105, Lys 106 and His 134). The movement of the loops will be discussed in Chapter 3. PPPK contains two Mg^{2+} ions which are both six-coordinated. Mg1 interacts with the α-phosphate and β-phosphate groups while Mg2 interacts with β-phosphate and the γ-phosphate group. The Mg^{2+} ions also interact with the oxygen atoms in the sidechains of Asp 95 and Asp 97 (Pf Asp 113, Asp 115). Both the Mg^{2+} ions also coordinate with water molecules. Blaszczyk *et al.* (2000) noted that Asp 95 and Asp 97 have to undergo significant conformational changes to interact with Mg^{2+} and thus magnesium binding-induced fit is evident.

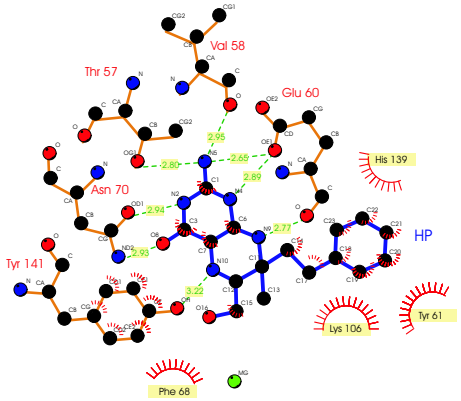
The *E.coli* structures also contained a substrate analogue (87Y) as well as the natural substrate HP (6-hydroxymethyl-7,8-dihydropterin). In the *E. coli* crystal structure HP undergoes ring stacking to stabilize the double ring system (Figure 2.16). Tyr 53 and Phe 123 are involved in ring stacking in the *E. coli* structure but in the PfPPPK model the situation appears to be reversed with Phe 68 fulfilling the role of Tyr 53, and Tyr 141 replacing Phe 123. HP also shows hydrogen bonds with *E. coli* residues Thr 42, Pro 43, Leu 45, Asn 55 (Pf Thr 57, Val 58, Glu 60, Asn 70) and these interactions saturate the hydrogen bond capacity of HP. Five other residues are also involved in binding HP: *E. coli* Gly 8, Tyr 53, Trp 89, Asp 95 and Phe 123 (Pf Gly 23, Phe 68, Lys 106, Asp 113, Tyr 141). Gly 8 is highly conserved between the different species and cannot be replaced by any other residue as this would interfere with HP binding through steric constraints, as well as preventing Asn 55 from changing its conformation to allow HP binding.

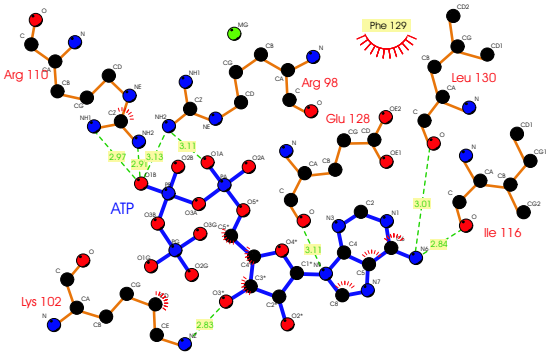


E. coli



H. influenza





P. falciparum

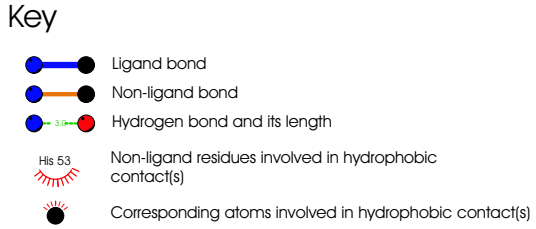


Figure 2.14: A comparison between the active site interactions of the *P. falciparum* PPPK model and known PPPK crystal structures. 87Y (*E. coli*) and Roi (*H. influenzae*) are the substrate analogues in the crystal structures. HP is the substrate analogue from the crystal structures built into the *P. falciparum* PPPK model. All figures were generated with LIGPLOT (Wallace *et al.*, 1995).

Table 2.2: Active site residue comparison between the crystal structures of *E. coli*, *H. influenzae* PPPK and the *P. falciparum* PPPK model. Some residues differ but the type of interaction between the residue and the substrate analogue remains the same. Note that most of the ATP-interacting residues interact with ATP though amide and/or carbonyl groups and thus the sidechain nature of these residues are relatively unimportant. Adapted from Blaszczyk *et al.* (2000).

E. coli	***H. influenzae***	***P. falciparum****	**Interaction with PPPK**
Gly 8	Gly 9	Gly 11/23	Interacts with HP, highly conserved
Thr 42	Ser 43	Thr 45/57	Forms hydrogen bond with primary amine on HP
Pro 43	Lys 44	Val 46/58	Use peptide backbone O for HP interaction
Leu 45	Leu 46	Glu 48/60	Glu 60 uses εO to form hydrogen bond with HP, Leu 45 uses backbone O for hydrogen bond formation
Tyr 53	Tyr 54	Phe 68/161	Sidechain associates with HP, substrate sandwiched between Tyr (Ec 53, Hi 54) and Phe (Ec F123, Hi F123)
Asn 55	Asn 56	Asn 70/163	Interacts with the 4-oxo group on ATP
Gln 74	Gln 76	Lys 78/185	Interacts with ATP through the NH_X groups of Lys and Gln
Glu 77	Glu 78	Glu 81/188	Interacts with Mg^{2+} and with ATP
Arg 82	Arg 83	Arg 86/193	Interacts with oxygens on the α-phosphate and the β-phosphate group of ATP
Arg 84	Arg 85	Asn 88/195	Interacts with oxygens on the α-phosphate group of ATP, helps with loop stabilization
Arg 88	Arg 88	Lys 93/200	Interacts with ATP through its ϵ-amino group
Trp 89	Trp 89	Lys 94/201	Interacts with substrate and γ-phosphate oxygen of ATP
Arg 92	Arg 92	Arg 98/205	Helps with loop stabilization and interacts with oxygens on the α-phosphate and the β-phosphate group of ATP
Asn 95	Asn 95	Asp 101/208	Interacts with HP and Mg^{2+} through its carboxy oxygens
Asn 97	Asn 97	Asp 103/210	Interacts with Mg^{2+} through its carboxy oxygens
Ile 98	Ile 98	Ile 104/211	Residues interact with ATP through their amide and/or carbonyl groups
Arg 110	Arg 110	Glu 116/313	Arg 110 use backbone O for hydrogen bond with the ribose part of ATP, Glu 128 uses terminal εO
Thr 112	Thr 112	Leu 118/315	Both residues uses peptide backbone O for interaction
Pro 114	Pro 114	Ile 120/317	Forms cavity for ATP, interacts with ATP core
His 115	His 114	Pro 121/318	Forms cavity for ATP, interacts with ATP core
Tyr 116	Tyr 116	His 122/319	Forms cavity for ATP, interacts with ATP core
Arg 121	Arg 121	His 139/324	Forms cavity for ATP, interacts with ATP core
Phe 123	Phe 123	Tyr 141/326	Sidechain associates with HP, substrate sandwiched between Tyr (Ec 53, Hi 54) and Phe (Ec F123, Hi F123)

* The first number indicates the residue in the model. The second number indicates the residue in the full bifunctional protein.

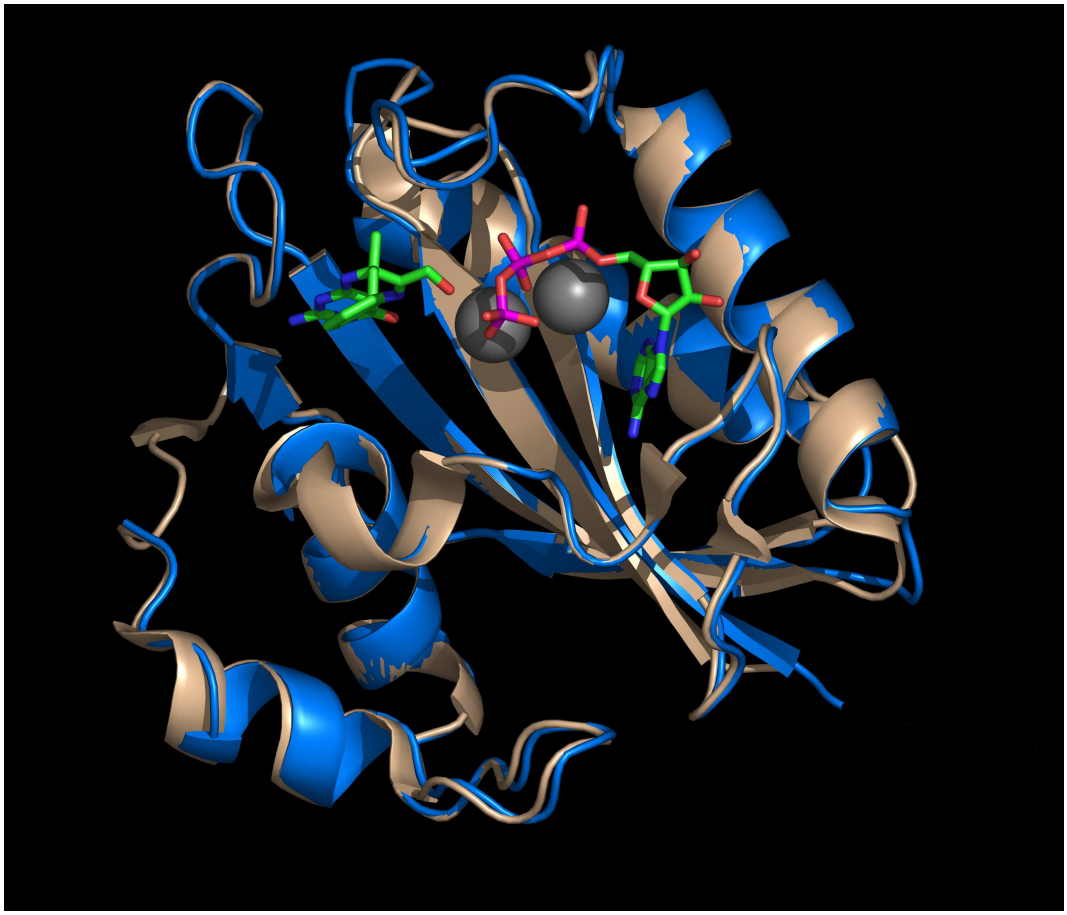


Figure 2.15: A comparison of *E. coli* PPPK (beige) with the PfPPPK model (blue). The model does not contain any of the parasite-specific insertions. The substrate analogue, 87Y is the stick model on the left and ATP is the stick model on the right. The Mg^{2+} ions are grey spheres. Note how the Mg^{2+} ions anchor the ATP. All the interactions are shown in Table 2.2. Generated with PYMOL (DeLano 2002, http://www.pymol.org).

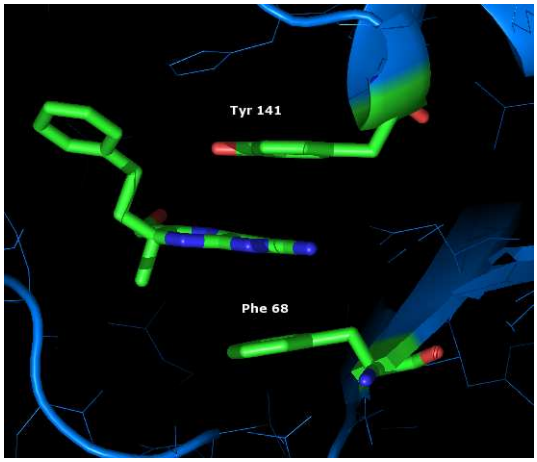


Figure 2.16: The ring stacking used to keep the substrate analogue in place in PPPK. *E. coli* uses Tyr 53 and Phe 123 while in the PfPPPK model the situation is reversed with Phe 68 and Tyr 141 being used. Generated with PYMOL (DeLano 2002, http://www.pymol.org).

2.4. Discussion

2.4.1. DHPS Homology Model

The DHPS model will be discussed in three sections, with the first section focusing on the homology model, the second section on the parasite-specific inserts and their effect on the PfDHPS model and the third section on ligand-protein interactions.

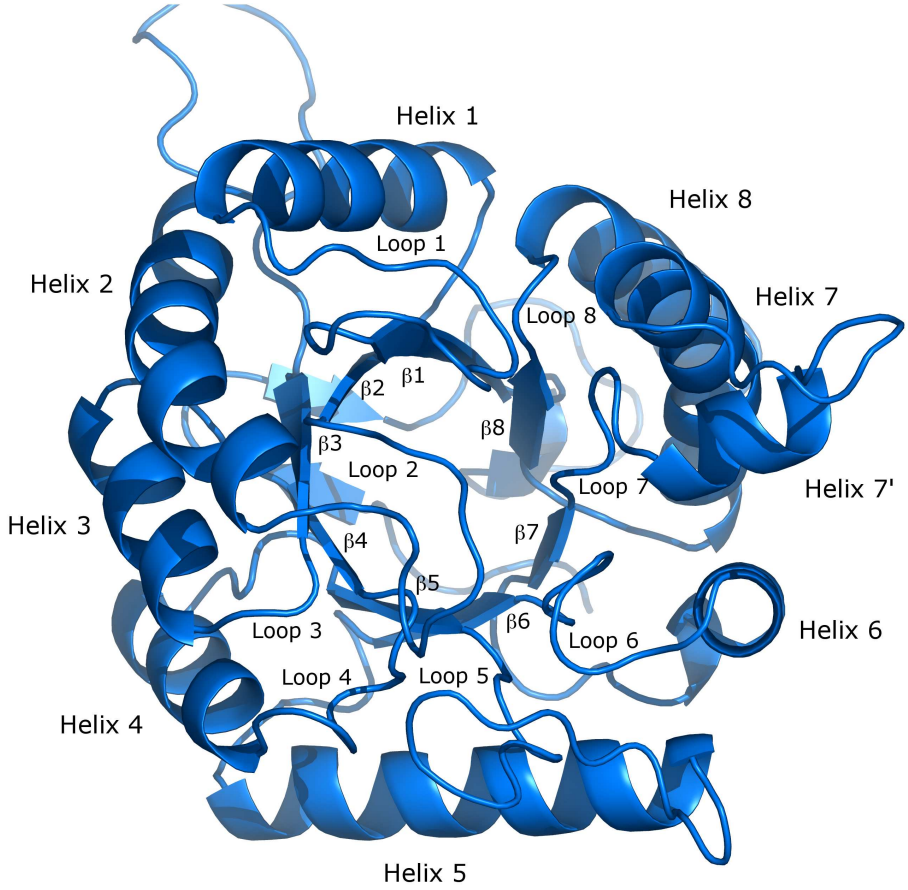


Figure 2.17: The modelled structure of *P. falciparum* DHPS. The model maintains the core barrel structure of *M. tuberculosis* (9 α-helices and 8 β-sheets). Helices 6, 7, 7' and 8 have been proposed to be involved in dimer interaction in *M. tuberculosis* and *S. aureus*, however only helix 7 and 8 are involved in dimer formation in *B. anthracis* (Babaoglu *et al.*, 2004). Insert 2 is located just after helix 7'.

2.4.1.1. *P. falciparum* DHPS model

The *P. falciparum* DHPS homology model compared well to the available crystal structures, especially in the active site region. The major structural difference between the model and the crystal structures is the inserts in the *P. falciparum* model. The annotated structure of *P. falciparum* DHPS is shown in Figure 2.17. The longest insert (insert 2, 39 residues) was left out in the construction of the model. No template is available and this region would thus only form an unordered peptide chain. In the crystal

structure of *B. anthracis* and *E. coli*, there is an anti-parallel β-sheet at the bottom of the TIM-barrel but not in the crystal structures of *S. aureus* and *M. tuberculosis*. The model contains an anti-parallel β-sheet at the bottom of the barrel and this is based on the *B. anthracis* template used. Support for or against the existence of this region as an anti-parallel β-sheet in *P. falciparum* DHPS could not be found. An interesting feature of DHPS structures in general is that the loops of the N-terminal end of the barrel sheets between the sheets and helices are very short (+/- 7 residues) while the C-terminal end has very long loop sections (some in excess of 30 residues). This correlates with the observation that the active site is located at the C-terminal end of the barrel sheets in TIM-barrel enzymes (Nagano *et al.*, 2002). Baca *et al.* (2000) and Hampele *et al.* (1997) speculated that loops 1, 2, 5 and 7 play a role in catalysis and that helices 6, 7, 7' and 8 play a role in protein-protein interactions in *M. tuberculosis*. The same loops are implicated in catalysis in *B. anthracis* but protein-protein interactions are mediated only by helices 7 and 8 (Babaoglu *et al.*, 2004).

2.4.1.2. *P. falciparum* DHPS-Specific Inserts

As seen from the alignment (Figure 2.4) *P. falciparum* DHPS contains two inserts in comparison to the template used for modelling. The high conservation of insert 2 between five *Plasmodium* species, implies the need to retain this part of the insert. The functional importance of inserts has been proven for *P. falciparum* dihydrofolate reductase-thymidilate synthase (DHFR-TS, Yuvaniyama *et al.*, 2003) and S-adenosylmethionine decarboxylase/ornithine decarboxylase (AdoMetDC/ODC, Birkholtz *et al.*, 2004). Yuvaniyama and co-workers have shown that the insert between the DHFR and the TS domains (containing two short α-helices) are involved in dimerization and domain organization. The helix in the insert serves as an anchor between the domains thus helping to keep the domains together. Birkholtz *et al.* (2004) deleted various combinations of parasite-specific inserts to determine their functionality. When certain inserts were deleted a 95% loss in enzyme activity was observed. When the connecting hinge region between the two domains were deleted an activity loss of 33% for AdoMetDC and a $\approx$50% loss in ODC activity was observed. When the domains were co-expressed with their respective inserts deleted, they also observed a marked loss in activity as well as a decrease in the formation of a functional, hetero-tetrameric enzyme. These results suggest the importance of inserts in domain-domain interactions as well as in enzyme activity.

The inserts in the known Plasmodial DHPS sequences differ with respect to their length and conservation. Insert 1 showed a difference in insert length between the species with *P. chabaudi* being the longest and *P. falciparum* the shortest. *P. chabaudi*, *P. berghei* and *P. yoelli yoelli* sequences show a high degree of conservation for insert 1 but the reason why it is not conserved between all five species is not known. A possible reason for this may be that these three species are more closely related to each other than

they are to *P. vivax* or *P. falciparum*. Kedzierski and co-workers (2002) used the adenylosuccinate lyase gene to do a phylogenetic analysis of the *Plasmodium* family. In this study they came to the conclusion that there are three main branches in the *Plasmodium* family. *P. berghei, P yoelli yoelli* and *P. chabaudi* were closely related on one branch while *P. vivax* and *P. falciparum* each occupied a separate branch. This supports the observation that the highly conserved sequences in insert 1 of *P. berghei, P. yoelli yoelli* and *P. chabaudi* are due to a recent common ancestor in the evolution of the *Plasmodium* family (Kedzierski *et al.*, 2002).

Insert 2 is interesting with respect to the secondary structure predicted for the last residues (747-759) in the *P. falciparum* sequence. The fact that the predicted helix is highly conserved between all five species may suggest that it has a functional role in the Plasmodial PPPK-DHPS enzyme. As seen from the work of Yuvaniyama *et al.* (2003), it may be involved in domain-domain interactions either between the PPPK-DHPS domains or between similar domains (DHPS-DHPS or PPPK-PPPK) and may also influence enzyme activity. Once again (as for insert 1) the highest sequence conservation for insert 2 is between *P. berghei, P. yoelli yoelli* and *P. chabaudi* but in this insert (in contrast with insert 1) *P. falciparum* and *P. vivax* show quite high sequence similarity with the other three species, indicating a need to conserve the sequence of insert 2. Interestingly, *P. vivax* and *P. yoelli yoelli* each contain species-specific inserts with the *P. yoelli yoelli* insert disrupting the predicted α-helix. Another interesting feature is the occurrence of a repetitive (G/A)KLTNG(D/E) ([Gly,Ala]-Lys-Leu-Thr-Asn-Gly-[Glu,Asp]) motif in the *P. vivax* specific insert. The [G,A]KLTNG[D,E] motif repeats nine times to form the *P. vivax* specific insert. The *P. yoelli yoelli* specific insert displays no such motif, and no repetitive motif was observed in any of the DHPS sequences from other *Plasmodium* species. The function of the repetitive elements remains unknown.

2.4.1.3. Ligand Interactions

The crystal structures of DHPS from *M. tuberculosis* (containing DHP-monophosphate), *B. anthracis* (various models containing pteroic acid, DHP-monophosphate, DHP and MANIC (6-methylamino-5-nitro-iso cytosine)), *E. coli* (sulfate ion) and *S. aureus* (DHP) revealed the conserved contacts made between the substrates and the protein. Table 2.1 lists the most important conserved residues. The interaction between the oxygen on the pyrazine ring and Lys is conserved between all the crystal structures (Pf 609, Mtb 213, Ba 220, Ec 221, Sa 203) as well as the interaction between the amino group on the pyrazine ring, Asn and Asp (Pf 503 and 575, Mtb 105 and 177, Ba 120 and 184, Ec 115 and 185, Sa 103 and 167). Asp also shows a conserved interaction with N8 of the pyrazine ring (Pf 482, Mtb 86, Ba 101, Ec 96, Sa 84). From the *S. aureus* and *B. anthracis* crystal structures the position and orientation of the pyrophosphate groups of DHP can be seen. The pyrazine ring structure also undergoes $\pi - \pi$-stacking

interactions with Arg (Pf 686, Mtb 253, Ba 254, Ec 255, Sa 239). It is well known that the pyrophosphate group interacts with metal ions. The position of the metal ion was determined from the structures of *S. aureus* (Mn^{2+}) and *M. tuberculosis* (Mg^{2+}). Kasekarn *et al.* (2004) determined that *P. falciparum* uses a Mg^{2+} ion for catalysis while *S. aureus* uses a Mn^{2+} ion. The *B. anthracis* and *M. tuberculosis* structures show that some of the oxygen atoms on the pyrophosphate group interact with the metal ion and with Arg (Pf 686, Mtb 253, Ba 254, Ec 255, Sa 239) as well as with His (Pf 688, Mtb 255, Ba 256, Ec 257, Sa 241). The metal ion is also anchored by the His residue (Pf 688, Mtb 255, Ba 256, Ec 257, Sa 241). This orientation of the pyrophosphate group exposes the carbon atom linking the pyrazine and pyrophosphate group for a SN_2 nucleophilic attack by the amino group of *p*-aminobenzoic acid (Baca *et al.*, 2000, Babaoglu *et al.*, 2004). The pyrophosphate group geometry may also create a partial positive charge on the linking carbon atom.

The structure of *B. anthracis* DHPS provided the only indication of where *p*-aminobenzoic acid is located in the active site. The *E. coli* structure has SAN (a *p*-aminobenzoic acid analogue) in the active site but it is in an unfavourable position to form pteroic acid. The *B. anthracis* structure indicated that the benzene ring of *p*-aminobenzoic acid undergoes hydrophobic interactions with the side chain of Lys (Pf 609, Mtb 213, Ba 220, Ec 221, Sa 203). The conformation of Lys 220 changes to accommodate the hydrophobic interaction with *p*-aminobenzoic acid when comparing the crystal structure with a ligand to the unbound structure. As a result of the *p*-aminobenzoic acid stacking with Lys, the carboxy group of *p*-aminobenzoic acid is in close enough proximity to form hydrogen bonds with the γ-O oxygen and the amide backbone nitrogen of Ser (Pf 607, Mtb 213, Ba 221, Ec 219, Sa 201).

As sulfadoxine is a *p*-aminobenzoic acid analogue it maintains most of the same interactions as *p*-aminobenzoic acid. The carboxyl group interactions of *p*-aminobenzoic acid are however replaced by the two oxygen atoms on the sulfur. This causes *p*-aminobenzoic acid and sulfadoxine to compete for the active site. Sulfadoxine may have some hydrophobic associations with the sidechain of Lys 540 through the methoxy groups. The K540E mutation shortens the sidechain with two methyl groups as well as reversing the charge. This may cause a loss of hydrophobic interaction with sulfadoxine. Sulfadoxine may also have some added interactions with Ala 437 although the orientation of Ala 437 did not indicate this. The S436F/A mutation causes a loss of the possible interaction between Ser 436 and sulfadoxine. The S436F mutation causes steric hindrance which prevents sulfadoxine from entering the active site. Ser 436 interacts with sulfadoxine through it's hydroxyl group, which breaks this interaction. Ala 581 shows polar interactions with the methoxy groups on sulfadoxine. The A581G mutation makes the chain more flexible, which may result in a reduced capacity to form polar interactions with sulfadoxine. The A581G does not seem to have a large effect on sulfadoxine binding and this may explain why it is one of the less common mutations in PfDHPS. The model could not explain the A613T/S mutation effects.

2.4.2. PPPK Homology Model

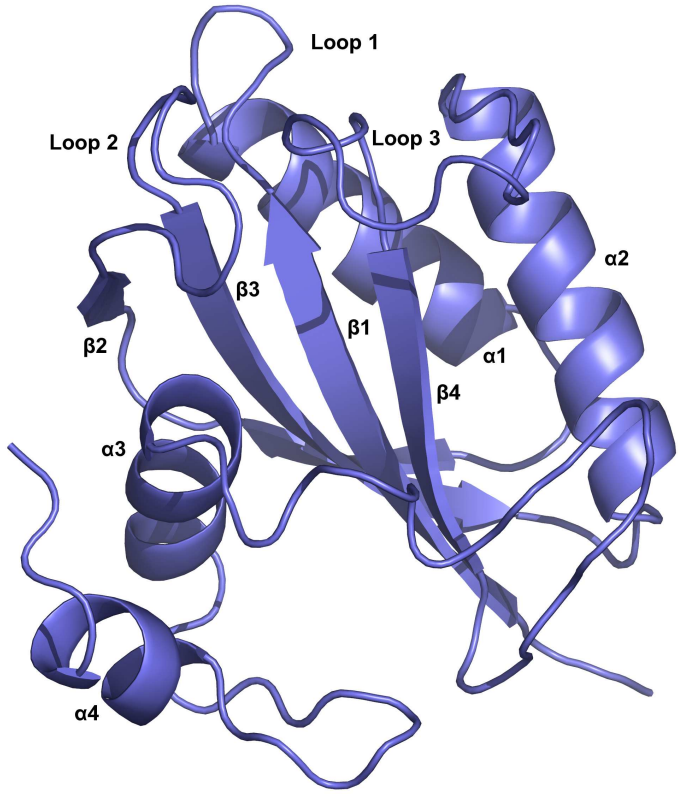


Figure 2.18: The annotated structure of the PfPPPK model. Insert 1 is located after β-sheet 2 and insert 2 is located just before β-sheet 4.

Figure 2.18 shows the annotated PfPPPK model with the loops, helices and sheets indicated (based on the annotation by Blaszczyk *et al.* (2000)). The movement of the loops will be discussed in Chapter 3.

2.4.2.1. PfPPPK inserts

The preliminary alignment showed that some residues aligned between the target and template, but when the model was built the residues did not occupy equivalent coordinates and could thus not fulfill their biological role. This required some of the residues in the alignment to be shifted to ensure that the model represented the active site accurately. When the sequences of the five *Plasmodium* species were compared with those of *E.coli* and *H. influenzae*, the presence of inserts in the *Plasmodium* sequences became apparent. Once again (as is the case with PfDHPS) it seems that *P. falciparum* PPPK contains two inserts in relation to other known species. The alignment (Figure 2.11) supports the hypothesis proposed for DHPS that *P. berghei*, *P. chabaudi* and *P. yoelli yoelli* are more closely related (Kedzierski *et al.*, 2002). From the alignment it is apparent that the second insert is much more conserved between *P. berghei*, *P. chabaudi* and *P. yoelli yoelli* than when compared to those of *P. falciparum* and *P. vivax*.

The three related Plasmodia also share a short insert in the middle of insert 1. In the *P. yoelli yoelli* sequence the first 30 residues do not match up to the other *Plamodium* species. *P. vivax* also seems to contain a short insert at the start of the PPPK sequence. The *P. falciparum* PPPK sequence contains an additional insert inside global insert 2. This is a low-complexity insert as indicated by SEG (Wootton *et al.*, 1993) and is a region rich in Asn residues.

2.4.2.2. Ligand interactions

PPPK contains two vital Mg^{2+} ions in the active site. Blaszczyk *et al.* (2000) suggests six possible functions of the Mg^{2+} ions in PPPK. Firstly, the ions may bring all the catalytic residues into the active site. Secondly, the ions orientate the hydroxyl oxygen of 6-hydroxymethyl-7,8-dihydro-pterin, the β-phosphorous and the oxygen between the α-phosphorous and the β-phosphorous of ATP. Thirdly, the ions may activate the β-phosphorous for a nucleophilic attack. Fourth, the ions lower the pKa of the hydroxyl group on 6-hydroxymethyl-7,8-dihydropterin and thus facilitate the reaction. Fifth, the Mg^{2+} ions can stabilize the transition state of the reaction. Sixth, the removal of AMP may be facilitated by the Mg^{2+} ions. PPPK also has various isozymes and Blaszczyk *et al.* (2000) found that Thr 42 was conserved among eight species but had various other amino acids in other isozymes. In this respect PfPPPK and BaPPPK are similar as they both contain a threonine at postion 42 (Table 3.1).

2.5. Conclusion

Worldwide resistance of the malaria parasite to sulfa-drugs is on the increase and new methods need to be investigated to identify new drugs and drug targets. In the previous sections models for *P. falciparum* DHPS and PPPK were presented. From the crystal structures (as well as the model) it was seen that the active sites of PPPK and DHPS must be able to bind the pterin moiety. Stammers *et al.* (1999) investigated the PPPK and DHPS active sites of *E. coli* to look for similar binding cavities. They found that the hydrogen bonding around the edges of the pterin moiety is the same between the two enzymes but that the phosphate-binding sites are different. In *E. coli,* DHPS interacts directly with the pyrophosphate groups through basic residues. In *E. coli* PPPK the pyrophosphate groups interact through cations (Mg^{2+} ions) as well as hydrogen bonds with basic residues. The same principle holds for the PPPK and DHPS models of *P. falciparum.* Using the pterin moiety as a scaffold, a drug could be designed which would inhibit both enzymes as these enzymes are not found in mammals. This double-pronged approach would alleviate some of the problems of drug resistance in DHPS enzymes.

Using molecular dynamics, the next chapter investigates the effect of resistance-causing mutations on the binding of the natural substrate to DHPS as well as on the binding of sulfadoxine to DHPS.

Chapter 3

Molecular Dynamics

3.1. Introduction

What is Molecular Dynamics (MD)? In simple terms (http://www.wikipedia.org):

> "Molecular Dynamics numerically solves Newton's equations of motion on an atomistic or similar model of a molecular system to obtain information about it's time-dependent properties".

This statement translates to the calculation and observation of the predicted motion of atoms in a molecule over certain time periods. Although crystallography can resolve the structure of a protein at a very high resolution (Jelsch *et al.*, 2000), it reflects only on rigid crystals and not the dynamic features of a protein in solution. The problem is that not all proteins can be easily crystallised or expressed in sufficient quantities. Nuclear Magnetic Resonance (NMR) can be used to solve the structure and motion of the protein but the protein is needed in soluble form and at high concentrations. This comes back to theproblem of protein expression and solubility (as is the case for membrane proteins). To use MD and Normal Mode Analysis (NMA), only a modelled structure is needed.

MD started out in theoretical physics but was applied to Biochemistry in the early 1970's. It has been applied to a vast range of protein motion problems such as the transfer of water across the aquaporin protein (de Groot *et al.*, 2001), folding of small proteins (Fersht, 2002), calculating the mechanical force exerted by G proteins (Kosztin *et al.*, 2002), elucidating the general mechanism of anesthesia (Tang *et al.*, 2002) and resolving the mechanism of the GroEL complex (Ma *et al.*, 2000). It has also been used to investigate the motion of loops in a protein, more specifically the triosephosphate isomerase enzyme (TIM) (Derreumaux *et al.*, 1998). Due to its theoretical nature MD has often been seen as a toy with which computational biologists play to make proteins move. However, MD can also provide thermodynamic properties, calculate protein structures at different temperatures and predict the effect of different solvents on protein structure. Two areas in which MD has been applied succesfully will be

discussed to answer problems which would have been difficult to solve with experimental work. Some principles of NMR, NMA and in-depth MD will be discussed.

3.1.1. Nuclear Magnetic Resonance (NMR)

NMR is the effect of nucleus-nucleus interactions emitting radiation when placed in an oscillating, electromagnetic field. NMR works by exciting the atoms with radio waves of a certain frequency and then measuring the emitted radiation. Only atoms with a nuclear moment of zero can be measured, including elements such as hydrogen, deuterium, carbon, nitrogen and phosphorous (Leach, 2001; Schlick, 2002). These atoms all have an uneven number of protons or neutrons. When the magnetic field is applied the spin of the nucleus changes and this generates a magnetic field, which is unique for each type of interaction. A hydrogen bonded to an oxygen will generate a field different to that of a hydrogen bonded to a nitrogen or even another hydrogen. By measuring all of these fields we can predict where certain atoms will lie. A program using amino acid templates will then try and match the templates to the measured positions of the atoms and predict a structure (Schlick, 2002). Because the proteins are in solution when doing NMR, different positions of the atoms can be measured resulting in a snap-shot series of protein movement. This is still not accurate enough because some atoms are in different environments than others, e.g. an atom in the centre of a protein will experience a hydrophobic environment while an atom on the surface of a protein will experience a water-rich environment and thus some magnetic fields will look different for the same atom. The nuclear Overhauser effect (NOE) can be used to enhance the resolution of NMR. NOE is the transfer of spin polarization amongst atomic nuclei and enhances the resonance of nuclei. This enhancement helps to identify more atoms when using NMR (Leach, 2001).

3.1.2. Normal Mode Analysis (NMA)

NMA is the analysis of motion in proteins using all the modes of movement. A protein will have different mode frequencies with each bond being a mode frequency. Thus an NMA analysis at a mode frequency of e.g. 12 000 will be an analysis of motion of every individual bond in a protein containing 12 000 atoms. NMA is usually done only for lower frequency modes up to a 100, but the most useful is mode frequency 7-11. These modes represent global protein movement and thus give a picture of the movement of different parts of the protein (Tama *et al.,* 2004; Brink *et al.,* 2004). This is a less time-consuming method than MD but does not give the same amount of information as an MD simulation does. With an MD simulation a very detailed analysis can be done on certain atoms in a system while with NMA it is more difficult to separate the individual atom motions from one another. NMA is also not time-dependent.

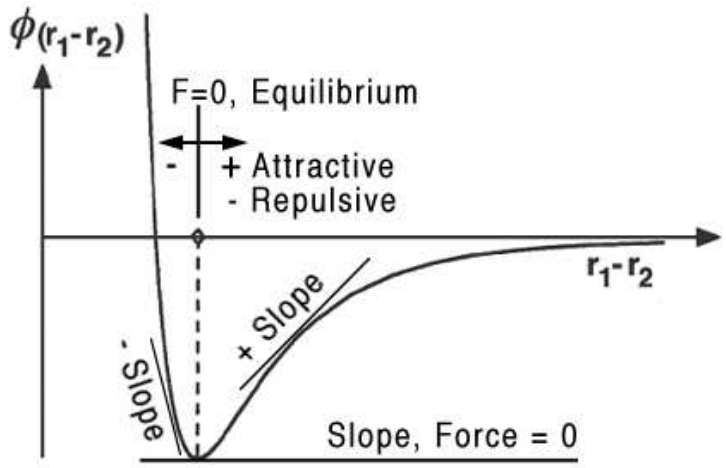


Figure 3.1: The Lennard-Jones potential. This function describes the interaction of two molecules or atoms over a distance. When the attractive and repulsive forces are equal a bond has formed or a stable interaction has occurred. As molecules move further away (r_1-r_2, indicated on the y-axis) from each other the energy goes to zero, but as they move closer the energy goes up as repulsion increases. F is the force experienced between the two molecules and ϕ is the energy.

3.1.3. Molecular Dynamics (MD)

To understand MD we need to grasp the underlying principles. A brief overview of these principles will be given followed by a discussion of how these principles are applied.

3.1.3.1. Forcefields

Forcefields (or molecular mechanics) are the basis of all molecular dynamics simulations. All interactions and movements of molecules during a MD run are based on forcefield parameters. Forcefields differ from quantum mechanics (QM) in that the electronic motion is ignored and only the energy of the system based on nuclear positions is calculated (Leach, 2001). Quantum mechanics is usually used for small molecules containing a few atoms. Forcefields can perform as well as quantum mechanic computations and this is mainly due to one assumption. This is the Born-Oppenheimer approximation, which states that since nuclear motion is much slower than electron motion the electronic wavefunction, or energies, can be calculated assuming a fixed position of the nuclei and nuclear motion can be considered assuming an average distribution of electron density. This enables us to calculate the energy of a system without doing any QM calculations on a system containing thousands of atoms (Schlick, 2002).

The Lennard-Jones potential (Figure 3.1) underlies all of MD and thus a thorough understanding of this concept is needed (Brooks *et al.*, 1983). The Lennard-Jones potential describes how the interaction energy changes when two molecules approach one another. Various modifications of this function can describe different interactions such as the simple interaction of two helium atoms or the complex interaction of two atoms forming a hydrogen bond.

There are four key contributions to the construction of a forcefield. These are bond stretching, angle

bending, torsional terms and non-bonded interactions. A forcefield usually contains energy penalties if these parameters deviate from the reference values. Some forcefields may contain terms for hydrogen bonds as well as extended terms for non-bonded interactions. A forcefield usually has two main parts, the potential energy ($\gamma(r)$) and the kinetic energy. The potential energy plays the main role in MD. The first derivative of the potential energy is related to the force acting on an atom and thus the movement of the latter. Potential energy is usually monitored closely during MD runs (Leach, 2001).

The potential energy function contains many terms or parameters such as bond potential, bond angles, torsions, cross-terms, non-bonded interactions, dispersion terms, repulsive terms, electrostatic terms and hydrogen bonds. The energy of the system is calculated by the following formula (Equation 3.1, Leach, 2001):

$$E = \frac{1}{2}\sum_{i=1}^{N_{df}} m_i v_i^2 + \gamma(r) \tag{3.1}$$

where γ(r) represents the potential energy term, mv represents the kinetic energy term.

$$E_{(total)} = E_b + E_\theta + E_\phi + E_\omega + E_{vdW} + E_{el} + E_{hb} + E_{cr} + E_{c\phi} \tag{3.2}$$

A forcefield like CHARMM (Brooks *et al.*, 1983) uses a more complex total energy function (Equation 3.2) where E_b is the bond potential, E_θ is the bond angle potential, E_ϕ is the dihedral angle torsion potential, E_ω is the improper torsions, E_{vdW} is the Van der Waals interaction, E_{el} is the electrostatic potential, E_{hb} is the hydrogen bonding, E_{el} is the atom harmonic constraint and $E_{c\phi}$ is the dihedral constraint.

The bond potential is the term which specifies how the energy changes when a bond between two atoms is stretched. The bonds will always deviate from the reference value due to vibrational energy at room temperature. Bond length changes are usually described by using a variation of Hooke's law (Equation 3.3):

$$v(l) = (k/2)(l - l_0)^2 \tag{3.3}$$

where l is the bond length, k is a constant and l_0 is the reference bond length (Leach, 2001; Schlick, 2002). The equilibrium bond length is the length when all the other parameters are set to zero. Thus, this bond represents an idealized state. The bond angle term specifies how the energy changes when the bond angle between three atoms is changed. Bond angle changes can also be described by Equation 3.3 but then l is substituted by θ, the bond angle. Each angle has a force constant associated with it. The force constant for bond angles are smaller than that for bond length since it requires less energy to change bond angles than bond lengths. The torsional terms describe how the energy changes as atoms rotate

around a bond. Because this only involves rotation around a bond and not actual modifying of bond characteristics these (together with non-bonded terms) are the terms contributing most to structural changes in systems. Torsion in molecules occurs when rotation around a bond causes atoms to undergo sterical clashes (McMurray, 2000).

Torsion potentials (improper torsion angles and potentials) can be used to maintain certain molecular characteristics like planarity in ring systems. A forcefield may do without a torsional term by simply relying on non-bonded interactions. Most forcefields for proteins and complex molecules do have an implicit torsional term. Torsions are usually expressed as a Cosine series expansion incorporating the torsion angle (ω) and a phase factor (γ) (Equation 3.4).

$$v(\omega) = \sum_{n=0}^{N} V_n/(2[1 + cos(n\omega - \gamma)]) \tag{3.4}$$

The phase factor describes the situation when a torsion angle goes through a minimum. Thus for ethane there will be three minima as rotation of the carbon-carbon bond results in the attached hydrogens interacting. The torsion angle at a minimum energy will thus be either 60°, 180° or 300° when starting with the staggered conformation. Modifying torsion terms on double bonds is almost the same as breaking a bond, thus torsion energies for double bonds will be much higher than normal torsional energies (McMurray, 2000). The non-bonded terms describe the interaction between atom pairs separated by at least three bonds but also atom pairs which are in different molecules. Non-bonded interactions may include hydrogen bonds, Van der Waals interactions and electrostatic interactions. These are usually represented as charges on the atomic nucleus or around the atom. It is very difficult to represent partial charges and there are a few methods to do this although none of them are currently based on experimental measurements of partial charges (Leach, 2001).

Hydrogen bonds may be represented as explicit hydrogen bond terms or as a result of Van der Waals and electrostatic interactions (Brooks *et al.,* 1983). Hydrogen bonds may also be represented by a modified Lennard-Jones potential (Equation 3.5). This function is used to predict the interactions between a hydrogen bond acceptor and a hydrogen atom.

$$v(r) = A/r^{12} - C/r^{10} \tag{3.5}$$

where A and C are parameters describing the distance at which collisions occur as well as the energy at which the collisions occur (Leach, 2001). Forcefields are primarily used to reproduce structural properties. Forcefields can also be parameterized to reproduce certain properties e.g. it should be possible to use the same set of parameters for all n-alkanes. Some parameters can also be transferred between forcefields without influencing the calculations too much. Parameters like bond angles and bond lengths can be readily transferred between forcefields. Most forcefields use the concept of atom types (Schlick, 2002).

Thus the parameters for a sp^3 carbon in a sugar and that for a sp^3 carbon in an alkane or alcohol will differ with respect to bond lengths and energies. This methodology is implemented in the CHARMM and AMBER forcefields, which both e.g. have different parameters for Histidine, depending on where it is protonated (Brookes *et al.,* 1983; Weiner *et al.,* 1984).

In order to save computational time the United Atom Force Field theory (UAFF) was developed. UAFF treats small molecules as single points rather than as molecules. A water molecule would not be represented as a three atom molecule but as a one point "atom". The same can be done e.g. for an ethylene molecule. UAFF will take the effects of the hydrogens into account when constructing the single "atom". This can result in a substantial decrease of computational time e.g. for butane there would be 144 terms to calculate with an atom based force field. Applying UAFF results in only 16 terms that need to be calculated. Even molecules such as benzene have been parameterized with UAFF (Toxvaerd, 1990).

In order to use non-standard molecules in forcefields, they need to be parameterized. This is the process of defining the bond lengths, bond angles and energies, dihedral angles as well as improper torsion angles for the molecule. Calculation of torsional energies is usually done via some quantum mechanical calculation. Although parameterization of forcefields for certain molecules can take up a lot of time, it can also be done by estimating the appropriate parameters based on similar molecules or bonds. This can produce reasonable results if the estimation is done using the right atom types.

3.1.4. Energy Minimization

Minimization basically refers to the energy surface being explored as a physical surface with valleys (corresponding to low energies) and peaks (which corresponds to higher energies). Minimization is the process by which we try to relieve the strain on a molecule and arrive at a global energy minimum (or valley) for a certain structure (it is almost impossible to know whether the local minimum is also the global energy minimum). The problem can be stated formally as:

> "Given a function f which depends on one or more independent variables x_1, x_2,...,x_i find the values of the variables for which f has a minimum value" (Leach, 2001).

There are two main methods of minimization, namely those which use derivatives and those which do not. In functions that uses derivatives the first derivative of a function will be zero at the minimum and thus indicate the minimum energy given a certain starting point. Most minimization methods cannot go uphill and thus all the methods will find the local minimum closest to the starting conformation.

In non-derivative methods of minimization there are two main methods namely the simplex method and the sequential univariate method (Leach, 2001). The simplex method makes use of a M+1 sided geometrical figure (where M is the dimensionality of the energy function), which is moved across the

energy surface. This is better than derivative methods but involves an immense amount of calculations and is thus often impractical. A few steps of simplex method followed by a more efficient method such as steepest descent has been shown to work well (Leach, 2001). The sequential univariate method is usually used for quantum mechanical calculations. It works by taking a coordinate, generating two new structures for that coordinate, calculating the energy and then fitting a parabola on the three points. The lowest point is then determined and the coordinate is changed to this lowest point. This process is repeated until no new lowest point can be found, indicating that the minimum has been reached (Schlick, 2002).

Derivative methods are divided into first order methods (which only use the first derivative) and second order methods (which use both first and second derivatives). Of the first order methods steepest descent (SD) and Conjugate gradient (CG) are the most widely used. With SD the first derivative is calculated as a vector, thus providing a direction for the next calculation step. When the minimum has been crossed the step size is decreased in order to converge on the local energy minimum. With each calculation the coordinates of the previous step are used and each successive step is at a right angle to the previous direction of movement. SD can result in small errors, especially when the minimum lies in a narrow valley and the step size is large enough to step over the valley. Thus the step size needs to be reduced when approaching the minimum.

In CG the gradient at each point is orthogonal but the directions are conjugate. A set of conjugate directions has the property that for a quadratic equation the minimum will be reached in an amount of steps equal to the amount of variables. One of the most widely used CG variations is that of Polak and Ribiere (Schlick, 2002; Leach, 2001).

The most commonly used second derivative method is the Newton-Raphson method, developed by Isaac Newton and later modifed by Raphson (Schlick, 2002). The first derivative is used to determine the minimum and the second derivative is used to determine the curvature of the function. Due to Hessian matrix (a n x n symmetric matrix of second derivatives) inversions being applied to calculate the energy surface, it is very computationally intensive and thus is used only for molecules less than 200 atoms in size. Variations have been introduced, which use the same inverted Hessian matrix for several steps but simply recalculates the gradients at each step.

Minimization methods will never reach the exact minimum (because of the step size) but will always get closer and closer, thus the need to be able to tell the procedure when to stop. The most common ways are to stop minimizing when the energy change between two successive steps reach a certain value, or after a certain number of steps. Another way is to stop minimizing when the change in coordinate movement is small enough or when certain RMS (Root-Mean-Square gradient) differences have been satisfied (Leach, 2001; Schlick, 2002).

3.1.5. Computer Simulation Methods

There are various computer simulation methods available to simulate molecular movement but the discussion will be limited to the two methods most used for proteins, namely Monte Carlo simulation and molecular dynamics.

3.1.5.1. Molecular Dynamics Simulations

Molecular Dynamics is the application of Newton's equation of motion to a set of points or atoms. These equations are solved for certain time intervals to obtain the movement of molecules over a certain time period. By using small time steps (in the femtosecond range) for each calculation, an accurate picture of how molecules move can be acquired. The coordinates of each point are adjusted after each calculation according to the new velocities and directions of movement determined by Newton's equations. The changes in position for each point can be combined into a trajectory, which describes the movement of each atom. These calculations have led to various insights into molecular movement such as the prediction of the motion of the loop in the TIM-barrel enzyme (Derreumaux *et al.,* 1998) and the subsequent confirmation by NMR (Rozovsky *et al.*, 2001). Derreumaux and co-workers predicted that the loop will close more slowly when there is no substrate present and Rozovsky and co-workers proved that the speed of loop opening is related to ligand availability and type of ligand. This is a good example of where MD can be used to predict properties, which can later be tested in the laboratory (Leach, 2001).

3.1.5.2. Monte Carlo Simulations

Monte Carlo (MC) simulation differs from MD in that it generates configurations randomly based on the previous conformation and then decides which configurations to keep, based on a set of criteria. Single atoms or even a few atoms may be moved at a time and the potential energy calculated. If the potential energy is lower than the previous state then the conformation is accepted. If the energy is higher than the previous state then the Boltzman factor of the energy difference is calculated. If the Boltzman factor is higher than a randomly generated number between 0 and 1, the conformation is kept else the original conformation is kept for the next round. This allows for a molecule to go uphill on the energy surface.

The difference between MD and MC is that MD can calculate time-dependant properties of a system whereas in MC the states cannot be linked through time. In MC the total energy is calculated with only the potential energy function whereas MD has a kinetic energy function as well (Schlick, 2002).

3.1.5.3. Periodic Boundaries

In order for molecules not to fly off into space during a simulation, certain boundaries need to be imposed on the system. There are two types of boundaries: periodic boundaries and non-periodic boundaries.

Periodic boundaries are used when molecules are investigated in solution, e.g. a solvated protein. A periodic boundary can be seen as an appropriately sized box (often a cube, rhombic dodecahedron or sphere), which is filled with solvent and solute. This box is then replicated in all three dimensions meaning that as soon as a solvent molecule leaves the current box another perfect copy of the molecule enters the box from the other side, thus always maintaining the same number of molecules in the box (Friedman, 1975). This is used to minimise the effect of boundary conditions on solvent interaction. When studying the adsorption of liquids onto surfaces, periodic boundary conditions can only be used on molecules which move parallel to the surface. Any molecule that strays out of the top of the box wil just be deflected back in. Periodic boundary conditions can be used with most shapes of boxes but a shape approximating a sphere will involve the least amount of calculations to be done (Leach, 2001).

Non-periodic boundary conditions are sometimes used in situations where a natural boundary such as in a water droplet occurs. It is also possible to only solvate a small part of the protein such as the active site although this may lead to errors when studying movement of molecules. When solvating only a part of the molecule, such as an active site, then a reservoir must also be set up in order to account for the effects of solvent molecules moving in and out of the central solvated zone. When only solvating a part of a molecule, errors may arise due to unnatural movement occurring when part of the complex is in solvent and another part of the complex is in a vacuum (Schlick, 2002).

3.1.5.4. Long-range Forces

When simulating the movement of molecules we cannot just consider the interactions between molecules, which are next to one another. We must also consider long-range interactions between e.g. residues in a protein. Usually some kind of cutoff is introduced where long-range interactions are considered zero after a certain distance. A group-based cutoff is used by assigning certain atoms in a molecule to groups. This has the effect that all interacting energies can be calculated on a group-group interaction basis thus using less time to calculate. In theory it is best to use no cutoff but this becomes too computationally expensive to use in practice. There are two kinds of group-based cutoffs: switched and shifted.

The shifted function modifies the Lennard-Jones potential until the cutoff point and then the interaction energies suddenly drop to zero. This will cause instabilities in the simulation and the potential will deviate from the "true" potential. Shifted functions work well with homogenous systems composed of one atom type, however this is disastrous when used in heterogenous systems (Schlick, 2002).

Switched functions use a continously-adjusting modification of the Lennard-Jones potential to calculate the interaction energies. The best version of the switch function is where it follows the Lennard-Jones potential until the cutoff and then changes to zero over a short distance. Thus the difference between the energies at the cutoff point is not as great as for the shifted function. Long-range forces can interact over very long distances and trying to compute this in a suitable solvent box is impractical. A few methods

have been developed to deal with long-range interactions without having to model all of the molecules involved. The Ewald summation method was developed to handle this problem. The Ewald summation method has been mostly used for highly polar solutions or solid-solid interactions but also increasingly for protein and DNA simulations (Ewald, 1921). The method uses Fast Fourier Transforms to calculate the electrostatic energy involved with the interactions. Instead of using discrete points for the charges, a Gaussian distribution is used to approximate the charges. Using Fast Fourier Transforms with a Gaussian distribution of charges, an accurate representation of the long-range Coulomb interactions can be obtained. The Ewald summation method is best used with periodic systems like crystals (whether protein or inorganic).

In the next section the application of these methods to the *P. falciparum* PPPK and DHPS homology models will be discussed.

3.2. Methods

Molecular dynamics requires a rigorous setup of the parameters involved in the simulation. General protein parameters have been defined but parameters for novel ligands/modified residues must be generated by the user. The aim of the molecular dynamics simulation of PfDHPS is to investigate four main factors: the orientation and movement of substrates in the active site, the movement of loops during a set time period, effects of resistance-causing mutations on substrate binding and the effect of the resistance-causing mutations on sulfadoxine binding. The movement of the substrate is important in determining the stability of the substrate, and hence the catalyzed reaction, in the active site. An integrated overview of DHPS and PPPK movement as well as mutation effects, can be gained by applying MD to the models.

3.2.1. *P. falciparum* DHPS Ligand Parameterization

The optimal way to parameterize ligands for a molecular dynamics simulation is to isolate the ligand and do infrared spectroscopy, NMR or X-ray studies to determine bond lengths and energies. This was unfortunately not feasible for this study as the main aim was not to parameterize the ligands. One computational approach is to use *ab initio* QM methods to parameterize the ligands. This method, however, is very time consuming. Another approach is to choose closely-related atom types, bond lengths and energies from the CHARMM parameter files and adjust them to suit the ligand (personal communication, Prof. L. Nilsson, CHARMM developer, Karolinska Institute, Sweden). In this case the major adjustments needed were in dihedral angle energies to force the rings in the ligands to stay planar.

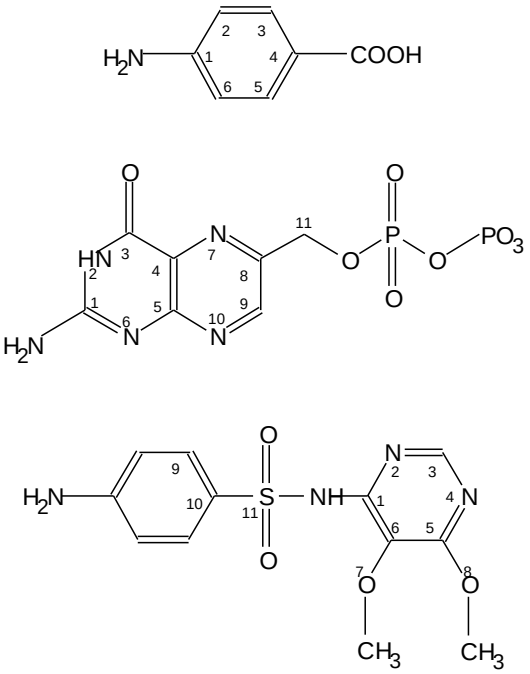


Figure 3.2: The structures of *p*-aminobenzoic acid (top), DHP (middle) and sulfadoxine (bottom). The three ligands were parameterized using closely related atom types and bonds from the the CHARMM parameter files. The dihedral angle energies were then adjusted to keep the appropriate bonds planar. Only the atoms involved in dihedral angle parameterization are numbered.

For the DHPS analysis three ligands required parameterization: *p*-aminobenzoic acid, DHP and sulfadoxine (Figure 3.2).

Several minimization interactions and parameter adjustments were needed to obtain the correct parameters for the ligands. Only the dihedral and improper torsion angles were modified during the parameterization process. The bond energy and bond lengths were kept at the default CHARMM values (MacKerell *et al.,* 1998). The numbers referred to at each ligand are the angles as indicated in Figure 3.2:

p-Aminobenzoic acid parameterization:

- Set the following dihedral angles at 0°: N-1-2-H; N-1-6-H; HOOC-4-3-H; HOOC-4-5-H.
- Set the following dihedral angles at 180°: N-1-2-3; N-1-2-5; HOOC-4-5-6; HOOC-4-3-2.
- Set the following improper torsion angles at 0°: H-C-C-H; N-C-C-H; C-C-C-H.

DHP parameterization:

- Set the following dihedral angles to 180°: H-2-3-O; H-2-1-6; 1-6-5-10.
- Set the following improper angles to 180°: O-3-2-1; 10-9-8-11; H-9-8-7; 9-10-5-6.

- Set the following improper angles to 0°: 4-5-6-1; 2-3-4-5; 7-8-9-10; 2-7-10-6; 5-10-9-H; 7-10-5-6; 4-5-6-1; 6-5-10-9.

Sulfadoxine parameterization (the *p*-aminobenzoic acid parameters were used for the analogous part of sulfadoxine):

- Set the following dihedral angles to 180°: N-1-6-5; N-1-2-3.
- Set the following dihedral angles to 0°: N-1-6-7; 2-3-4-5; 7-6-5-8.
- Set the following improper torsion angles to 0°: 1-6-5-7; 4-5-6-8; N-1-2-6.

The ligands were originally built in InsightII (Section 2.2.2.1 and 2.2.2.3) and docked into PfDHPS (Section 2.2.2.2 and 2.2.2.4). The ligands were subsequently parameterized for use in CHARMM during molecular dynamics.

3.2.2. Molecular Dynamics on *P. falciparum* DHPS

To compare the different states of PfDHPS (with substrate, without substrate, with different combinations of resistance-causing mutations) the starting state is needed. During the study of the dynamics of PfDHPS the same starting coordinates were used for all the different investigations. The protein and each of the respective ligands were merged using CHARMM to form one structure. The protein-ligand complex was then minimized for 500 steps using the Steepest Descent method. The Van Der Waals interactions were accounted for by using the vshift function of CHARMM. Vshift is a CHARMM method to account for long range forces. The non-bonded interactions used a cutoff of 12Å and were set to zero at 14Å. The protein-ligand complex was then solvated using the TIP3 water model and the CHARMM script supplied by Prof. L. Nilsson (personal communication, available at http://www.charmm.org). The script uses a box of 216 water molecules to fill the space specified around the protein. For PfDHPS it was decided to make a sphere of water (radius of 38Å) around the protein (protein diameter roughly 66Å). The protein-ligand complex was fixed and the water molecules subjected to minimization (50 steps SD followed by 50 steps ABNR minimization, Brooks *et al.*, 1983) to remove energy clashes between water molecules. Following water minimization the solvated protein-ligand complex was then minimized again (50 steps SD followed by 50 steps adopted basis Newton-Raphson (ABNR) minimization) to remove energy clashes between water molecules and the protein complex. The water molecules were contained in the 38Å radius sphere using a quartic sphere potential from the MMFP function in CHARMM. This

supplies a repulsive potential at 38Å to prevent the water molecules from moving off into space. The water molecules were restrained with a force of 10.0 kcal/mol. The molecular dynamics simulation was then started for differing lengths of time, depending on the characteristic under investigation. Each time-step was 0.001 ps long and the coordinates of the solvated protein-ligand complex were written out every 100 steps. To account for temperature the simulation was started at 0 K and increased in 10 K increments every 100 steps until 310 K was reached. The temperature was checked at each step to ensure that it did not increase too fast. Any unscheduled increase in temperature was normalized according to a Gaussian distribution at the current temperature.

3.2.2.1. Investigating Substrate Orientation and Movement

The solvated PfDHPS complex was used to investigate the orientation and movement of the substrate in the active site of PfDHPS. The movement of the substrate is important in determining the stability of the substrate, and hence the catalyzed reaction, in the active site. If the substrate does not remain in the active site it is not an optimal fit for the substrate. This simulation also serves as an indication of the correctness of the active site. The complex was subjected to a dynamics simulation for 100 000 steps (100 ps). For this investigation PfDHPS and the natural substrates (*p*-aminobenzoic acid, DHP and Mg^{2+}) were used. DHP and Mg^{2+} were started in the modelled position and *p*-aminobenzoic acid in the position indicated by the *B. anthracis* crystal structure.

3.2.2.2. Investigating Loop Movement

The model previously indicated that PfDHPS belongs to the TIM barrel family. Molecular dynamics (Derremaux *et al.*, 1998) studies of this enzyme have suggested that some loops move during catalysis and this has been proven using NMR and X-ray crystallography (Rozovsky *et al.*, 2001). Rozovksy and co-workers also speculated that the loop motion might be the reason why this enzyme is usually a rate-limiting step (at least partially) in some pathways. For this study the solvated and non-solvated PfDHPS complexes with the natural substrates in the active site were used. The simulation was run for 1 000 000 steps (1 ns) for each of the complexes on 6 nodes of a 64 CPU Linux cluster (each node: 2.4 GHz Pentium 4, 512MB RAM with Gigabit Ethernet, running SuSE Linux 9.1 and CHARMM 29b1).

3.2.2.3. Investigating Mutation Effects on Substrate Binding

Resistance to sulfa-drugs in PfDHPS is due to mutations in the enzyme. To investigate the effects of the different mutations, the solvated PfDHPS complex with the natural substrates in the active site was used. Only the A437G mutation was studied for its possible effect on natural substrate binding as all of the known resistance-causing mutations have no major effect on natural substrate binding (Triglia *et al.*, 1997). The simulation was run for 100 000 steps (100 ps).

3.2.2.4. Investigating Mutation Effects on Sulfadoxine Binding

As the mutations in PfDHPS only have an effect on sulfa-drug binding, the effect of the following resistance-causing mutations on sulfadoxine binding were investigated:

- S436F
- A437G
- K540E
- A581G
- A613T

Each simulation was run with the solvated PfDHPS complex containing Mg^{2+}, DHP and sulfadoxine in the active site. The effect of the mutations was compared to a simulation of sulfadoxine binding to wild-type, sulfadoxine-sensitive PfDHPS. Each simulation was run for 100 000 steps (100 ps).

3.2.3. Molecular Dynamics on *P. falciparum* PPPK

In order to assess the stability of PfPPPK and the substrate orientation and movement in the active site, the same simulation protocol as that for PfDHPS was used (section 3.2.2). A simulation time of 1ns (1 000 000 steps) was used for this study.

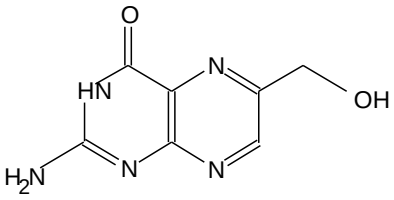


Figure 3.3: The substrate of PfPPPK, 6-hydroxymethyl-7,8-dihydropterin. The OH group is replaced by a pyrophosphate group from the ATP to form DHP (Blaszczyk *et al.*, 2000).

No additional parameterization of the PPPK substrate (Figure 3.3) was needed as CHARMM already contained parameters for ATP and the parameters for DHP were transferred to the substrate, 6-hydroxymethyl-7,8-dihydropterin. The parameters for the hydroxyl group on the substrate were added from the CHARMM parameter files.

3.3. Molecular Dynamics Results

Molecular dynamics is an attempt at simulating the natural world and thus cannot be taken as fact. Although it is theoretical we can still predict certain properties with great confidence and therein lies the power of molecular dynamics. In the next two sections the results of the molecular dynamics simulations of PfDHPS and PfPPPK are highlighted.

3.3.1. Substrate orientation and movement in *P. falciparum* DHPS

One of the aims of the molecular dynamics simulations was to test the stability of the protein model in a simulated water environment. All of the simulations showed that the PfDHPS model was stable in water. During the first few steps (10-20) of the simulation a relaxation of the protein structure could be seen. This is due to the difference in parameters used for model building and dynamic simulations. In a protein crystal structure a highly packed conformation implies that residue-residue interaction distances are closer than in the native, solvated state. These shorter interaction distances are transferred directly to the model during the process of homology modelling. As minimization only tries to resolve local energy clashes, it does not correct for the shorter interaction distances. The relaxation of the protein at the beginning of the simulation eliminates the effect of the crystal packing on the interaction distances. After the first few steps the PfDHPS structure stayed stable for up to 1 ns (longest length of simulation) in water. The active site also maintained it's conformation with the Mg^{2+} ion interacting stably with the protein. The stable interaction of the Mg^{2+} ion also provided a stable interaction for DHP. The DHP-Mg^{2+} interaction remained stable throughout all of the simulations conducted.

p-Aminobenzoic acid appeared to be mobile in the active site during the simulations. The simulations showed that *p*-aminobenzoic acid maintains an interaction with DHP for a few tens of picoseconds but after that the effect of the water breaks the interactions between *p*-aminobenzoic acid and DHP. This is partially due to the high solvent accessibility of the *p*-aminobenzoic acid. DHP shows very little movement during the simulation with most of the movement resulting from the different active site conformations sampled during the simulation. The interactions between DHP and the active site residues remain stable during the simulations, thus supporting the accuracy of the active site of the model. The Mg^{2+} remains stably anchored between DHP and the protein. The Mg^{2+} ion is mainly anchored by Asn 396 (absolutely conserved in all species, same as in *M. tuberculosis*, *B. anthracis*; in *S. aureus* the conserved Asn anchors a Mn^{2+} ion) and Asn 296 helps with coordination of the Mg^{2+} in the PfDHPS model. The interaction beween Mg^{2+} and DHP is mediated through the anchoring of the DHP pyrophosphate group by Mg^{2+}. The pyrophosphate groups interact with water molecules, which stabilise the negative charges on the phosphates as well as interacting with the charged residues in the active site through water-mediated bonds (Baca *et al.*, 2000).

From the *B. anthracis* crystal structure it was seen that *p*-aminobenzoic acid associates with the hydrophobic side chain of Pf Lys 609 (Babaoglu *et al.,* 2004). This was the first evidence for the location of *p*-aminobenzoic acid in the active site of DHPS.

3.3.2. Loop Movement in *P. falciparum* DHPS

To investigate the loop movement unsolvated and solvated complexes were chosen as well as solvated substrate-containing and non-substrate containing complexes. In the unsolvated complex the charges on

the protein are not screened by solvent and thus exaggerated and accelerated loop movements were seen. The unsolvated simulation showed that most of the loops at the opening of the barrel move closer to one another, eventually closing off the active site. The combined movement of the loops at the opening of the barrel screens the active site, and thus catalysis, from the interaction of the solvent. The solvated model indicated that loop movement does occur but on a smaller scale. This can be attributed to less interaction between the loops and more interaction between the loops and solvent. The unsolvated simulation suggested that the active site can be completely closed off from solvent and that the natural substrates stay in close enough proximity (less than 4 Å) to ensure catalysis. The solvated simulation showed that only partial closing-off of the active site occurs as there is some interaction between the solvent and some of the charged residues at the opening of the active site.

Loop 2 in *B. anthracis* has a bent conformation and this ensures that Arg 68 (Pf Ala 437) sticks into the TIM barrel and undergoes $\pi - \pi$-stacking interactions with Arg 254 (Pf Arg 686), which lies in a highly conserved region of DHPS. This $\pi - \pi$-stacking interaction between Arg residues is missing in *P. falciparum*. Loop 2 also has other stabilizing interactions between Lys 104 and Val 74 (Pf Asn 485, Lys 445), as well as between Arg 68 and Asp 101 (Pf Ala 437, Asp 482, interaction missing in *P. falciparum*) and a water bridge between the backbone amide of Gly 70 and Asn 27 (Pf Phe 439, Asn 396). The water bridge seen in *B. anthracis* between Phe 439 and Asn 396 seems to missing in *P. falciparum* as the two residues are too far apart (Figure 3.4).

From the crystal structures of *B. anthracis* (1TWS, 1TWW, 1TWZ, 1TX0 and 1TX2) is was seen that loops 3-8 have the same conformation in all the structures. The structures of *M. tuberculosis*, *S. aureus* and *E. coli* show that the orientation of loops 3-5 and 8 is not conserved and varies in length between species. Loop 5 shows movement during simulations (Figure 3.5) and this may be significant as loop 5 in *P. falciparum* contains Lys 540, which when mutated to Glu, causes resistance to sulfadoxine. Loop 6 forms part of the pterin-binding pocket and is highly conserved. The first few residues in loop 7 are highly conserved as they are close to the active site but the rest of the residues in loop 7 vary between species. The simulation showed that loop 7 is relatively flexible (Figure 3.6) and this flexibility, as well as the low sequence conservation, indicates that this loop might not play as an important role (although helix 7 is involved in protein-protein interaction) (Baca *et al.*, 2000; Babaoglu *et al.*, 2004).

3.3.3. Ala 437 Gly Mutation Effects on Substrate Binding in *P. falciparum* DHPS

The protocol specified in section 3.2.2.3 was used to investigate the effects of the A437G resistance-causing mutation on natural substrate binding. The simulations showed that the mutation does not have any significant effect on the binding of the substrates, although *p*-aminobenzoic acid did show some extra movement in some of the mutations but this was not deemed to be significant.

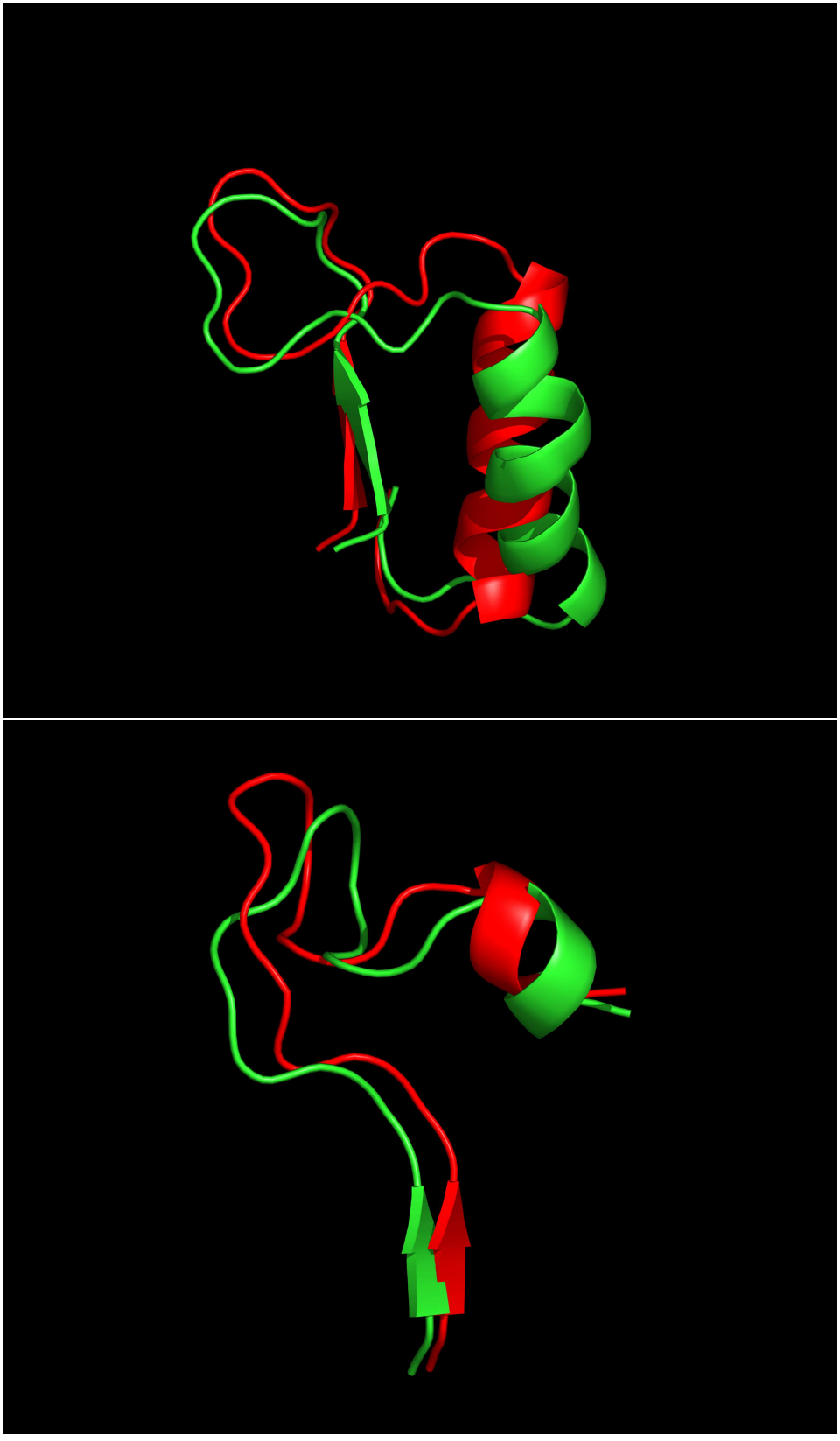


Figure 3.4: Refer to Figure 3.9 for location of various loops. Top: The movement of loop 1 in the PFDHPS model. Bottom: The movement of loop 2 during the simulation. Loop 2 anchors Mg^{2+} and although there is not much movement of the loop during the simulation, the helices move substantially while still maintaining the structured loop orientation. Loop 2 moves more in comparison to the limited movement of loop 1. Loop 2 contains Ala 437 and Ser 436 which, when mutated, cause resistance. The β-sheets in the figures are located on the inside of the TIM-barrel and the α-helices on the outside. Generated with PYMOL (DeLano, 2002).

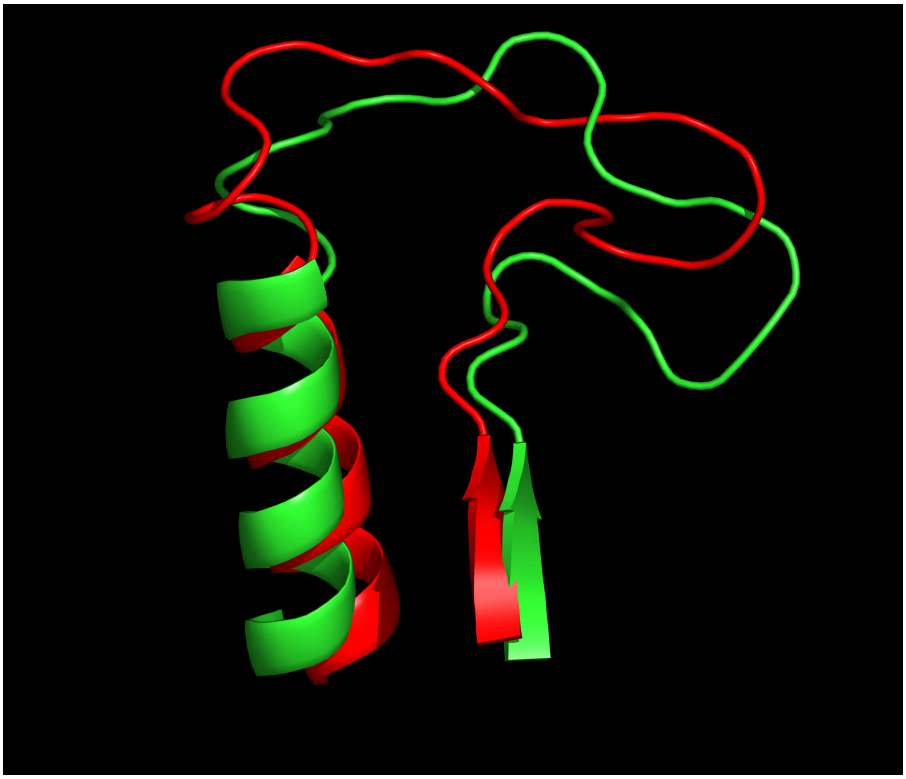


Figure 3.5: The range of movement of loop 5. Note the stability of the secondary structural elements. Generated with PYMOL (DeLano, 2002).

Figure 3.6: The range of movement of loop 7. Although loop 7 undergoes large conformational changes, these changes does not seem to be significant as loop 7 does not contain conserved residues nor does it interact with the substrate. Generated with PYMOL (DeLano, 2002).

3.3.4. Mutation Effects on Sulfadoxine Binding in *P. falciparum* DHPS

Resistance to sulfadoxine in PfDHPS is caused by mutations in the enzyme (Triglia *et al.*, 1997). The model indicated that most of the mutations are located near the active site but not in the active site directly (on loops 2 and 5). It was thus obvious that these mutations somehow affect the effectivity of sulfadoxine and that simulations of the different mutants might provide a clue to the mechanism behind resistance. The first simulation was done on the wild-type solvated complex with sulfadoxine. From the model it could be seen that the methoxy groups of sulfadoxine project into a relatively hydrophobic area containing, amongst others, residues 540 and 581. In the wild-type, sulfadoxine showed hydrophobic interaction with the protonated Lys 540 sidechain through the methoxy groups on sulfadoxine. In the K540E mutation simulation, the interaction with sulfadoxine is broken and the methoxy groups move towards the more hydrophobic residues. Glu 540 also associates more with the solvent, as well as forcing the loop to open up more than in the wild type and the other mutants (Figure 3.7). In the wild-type model, sulfadoxine shows hydrophobic interactions with Ala 581. The A581G simulation showed that the methoxy groups move away from residue 581 due to a decrease in hydrophobic interaction. In the S436F mutation the possible polar interaction between Ser 436 and sulfadoxine is lost. The Phe substitution also results in steric hindrance which may prevent sulfadoxine from binding. The current model could not explain the A613S/T mutation.

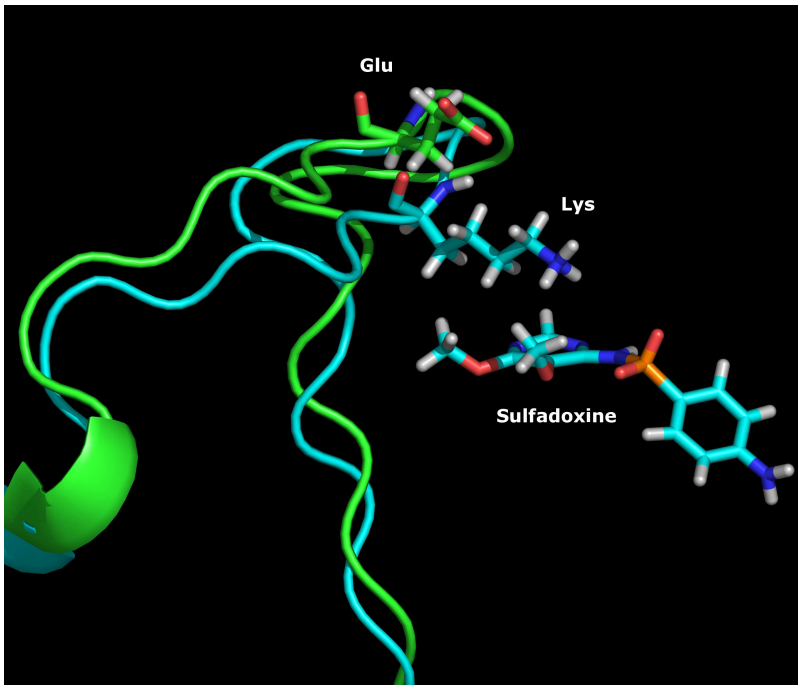


Figure 3.7: The effect of the K540E mutation on sulfadoxine binding. In the wild-type (light blue) the methoxy groups on sulfadoxine show interaction with the sidechain of Lys 540. Although the side chain is protonated there is no interaction between it and the sulfone group on sulfadoxine. When the K540E (green) mutation occurs, the methoxy groups rotate to point away (not shown) and the distance between residue 540 and sulfadoxine doubles. This may be due to the increased interaction of Glu 540 with the solvent, which may also explain why the loop opens up more than in the wild-type. Generated with PYMOL (DeLano, 2002).

3.3.5. Molecular Dynamics of *P. falciparum* PPPK

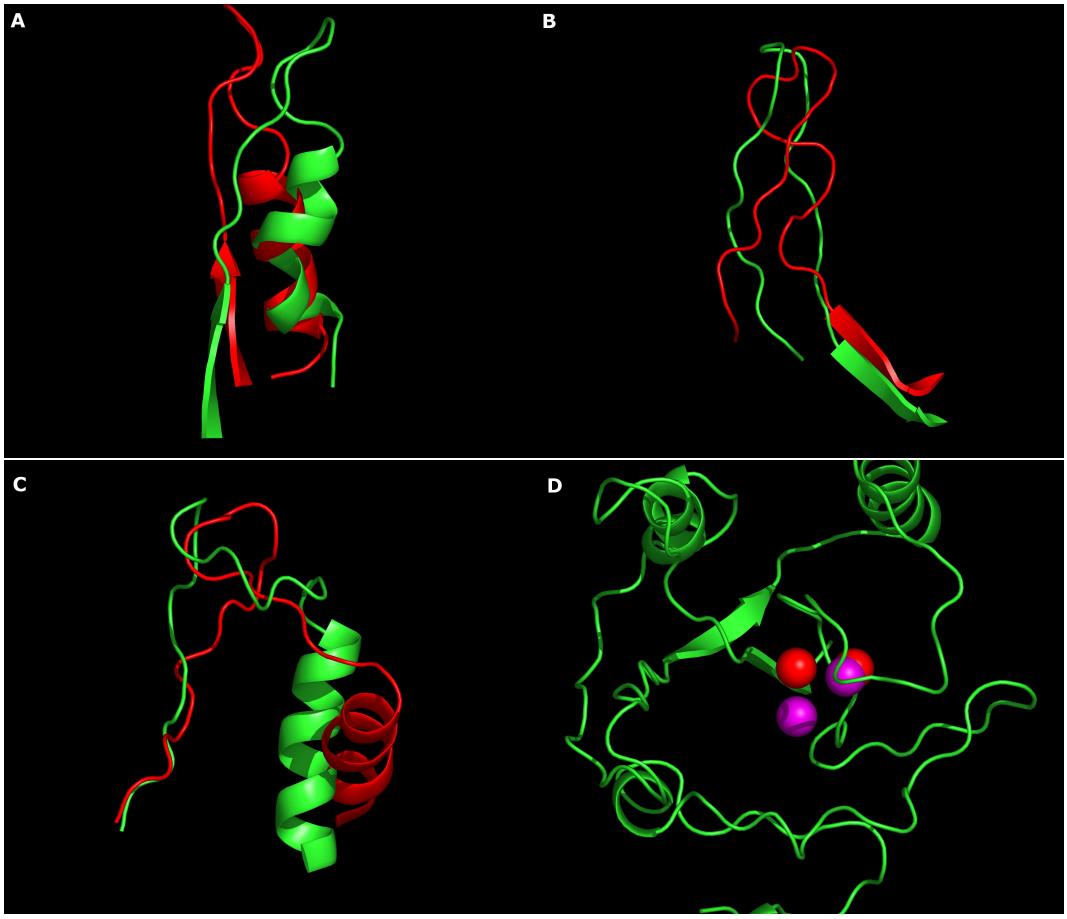


Figure 3.8: The movement of the different loops and the Mg^{2+} ions in the PfPPPK model. The starting position of the protein is indicated in green with red showing the position after the simulation. A: the movement of loop 1. Note the deformation of α-helix 1. B: The movement of loop 2. Once again a secondary structural element undergoes deformation with the end of the β-sheet deforming. C: The movement of loop 3 during the simulation. This figure shows the large conformational change of α-helix 2 during the simulation. D: The movement of the Mg^{2+} ions during the simulation. The substrates stayed in place during the most of the simulation indicating that the interacting residues maintained their conformations. The purple spheres represent the starting position of the ions and the red spheres the end postion. Generated with PYMOL (DeLano, 2002).

In *P. falciparum* PPPK forms a bifunctional enzyme with DHPS. This implies that PPPK and DHPS may interact with one another as shown for DHFR-TS (Yuvaniyama *et al.*, 2003) and AdoMetDC/ODC (Birkholtz *et al.*, 2004). The results of the molecular dynamics simulation of the PfPPPK model suggested that the model seems to rely on stabilization from DHPS in the bifunctional form. During the simulation some of the secondary structural elements became disordered (two β-sheets became disordered and one α-helix lost its conformation) and the enzyme undergoes an increase in its apparent size. β-sheets 2 and 4 seemed to lose their conformation as well as α-helix 3. Helix 2 seemed to undergo a $\approx$80°change in orientation which resulted in loop 3 showing some large movements (Figure 3.8). Despite this apparent instability of PfPPPK some loop movement could still be observed. Loop 3 associates with ATP and protects it from solvent attack but during the simulation loop 3 moved $\approx$11Å away and exposed ATP to

the solvent. During the simulation ATP stayed anchored in place by the Mg^{2+} ions but the substrate analogue moved ≈4.5 Å closer to ATP. The second Mg^{2+} ion moved 4.17 Å when the substrate analogue moved closer. Loop 2 moved ≈6Å and loop 1 moved ≈10Å from the starting position. The movement of the loops indicate that PPPK has extensive loop movement during catalysis as shown by Blaszczyk *et al.* (2000) and Yan *et al.* (2001).

3.4. Discussion

3.4.1. *P. falciparum* DHPS

3.4.1.1. DHPS Loop Movement

Derremaux *et al.* (1998) showed that TIM-barrel proteins have loops that move and close off the active site. As PfDHPS is a TIM-barrel protein it was decided to investigate the movement of the loops. The two main reasons for observing less loop movement in the solvated PfDHPS complex than the unsolvated complex are that the charges on the residues are screened by water, and that the substrates help to stabilize the loops and thus minimize loop movement. Another contributing factor may be that loop 1 and 2 are not in their proper conformation for movement. Loop 2 is missing from the crystal structure templates (the modelled structure of PfDHPS can be seen in Figure 3.9). Loop 1 and 2 should not be confused with insertion 1 and 2, as these were left out of the modelling process.

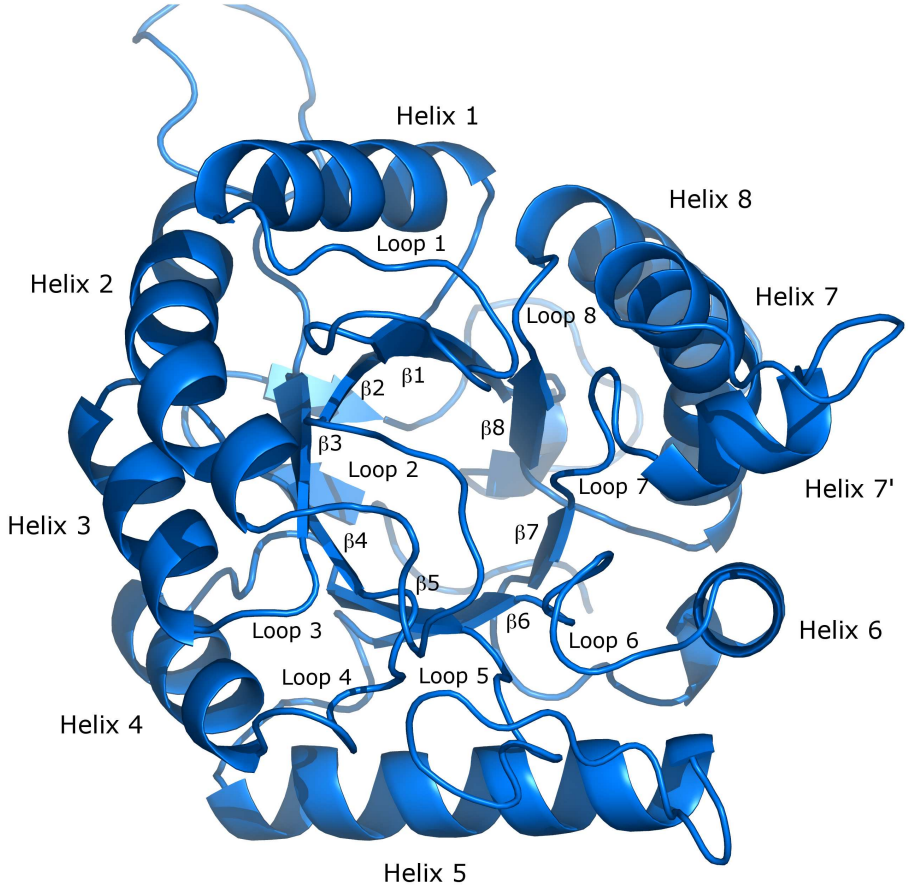


Figure 3.9: The modelled structure of *P. falciparum* DHPS. Insert 2 is situated just after helix 7'.

Baca *et al.* (2000) mentions four loops involved in catalytic function in *M. tuberculosis*. They failed to resolve loop 2 in the X-ray structure, leading to the conclusion that it is very mobile. The same loop is not visible in the *S. aureus* structure either (implying they are very mobile) and this supports the claim that the loops at the opening of the barrel are involved in active site closure. The *B. anthracis* crystal structure is also missing loop 2 as in *S. aureus* and *M. tuberculosis*. The loops also undergo interaction between them which helps to stabilize the closed conformation of the enzyme.

Some of the loops are also stabilized through interactions with certain residues. White *et al.* (2004) speculated that the loops are anchored in the following ways: Loop 1 (highly conserved) has a salt bridge with Arg 82, Glu 41 and Glu 65 (Pf Leu 451, Glu 410, Glu 434) as well as hydrogen-bond interactions between the side chains of Asp 31, Ser 34 and Ser 38 (Pf Asp 400, Ser 403, Ile 407). In *M. tuberculosis* loop 1 has an ordered structure (Figure 3.4) and Baca *et al.* (2000) proposed that Asp 21 (Pf Asp 45, Ba 35) is in contact with the substrate. There is a large difference in the orientations of loop 1 in *B. anthracis* and *M. tuberculosis*. In *M. tuberculosis,* loop 1 serves to anchor the Mg^{2+}. In *B. anthracis* it is too far away but may serve in identifying loop stabilizing residues or it may be a crystallization artifact. This may be due to the fact that *B. anthracis* DHPS was not crystallized with a Mg^{2+} in place.

Loop 2 in *M. tuberculosis* could not be resolved but the structures of *S. aureus* and *E. coli* showed the loop. In *E. coli* and *S. aureus* loop 2 has roughly the same orientation. The *B. anthracis* structure also revealed two water molecules which serve as bridges between Arg 68 and the side chains of Asn 120 and Asp 184 (Pf Asn 502, Asp 575), both residues which lie once again, in highly conserved regions. Loop 2 also forms a salt bridge between Arg 82 and Glu 65 (Pf Leu 451, Glu 434). In *P. falciparum* it seems that this salt bridge is absent. In the structures of *E. coli* and *S. aureus* it has been reported that loop 5 undergoes interaction with loop 2 (*E. coli* has extensive interactions between loop 2 and 5; *S. aureus* has a weak interaction between loop 2 and 5). This is in contrast with *M. tuberculosis* and *B. anthracis* where loop 2 and 5 show minimal interaction.

Some of the residues in the loops (residues 401-406 in the alignment in Figure 2.4) are highly conserved, thus indicating that they play an important function in the mechanism of DHPS. In the loop of *M. tuberculosis* Arg 54 (Pf Ala 437) is conserved as well as in *E. coli*, *S. aureus* and *M. tuberculosis* DHPS. Baca *et al.* (2000) suggested that this arginine is involved in binding the substrate through the reaction pathway. In all the *Plasmodium* species this arginine is replaced by Ala 437. This alanine is the residue which, when mutated to glycine, causes sulfa-drug resistance in *Plasmodium* species. Baca *et al.* (2000) speculated that some of the loops (loops 2, 5, 6 and 7') are involved in forming some kind of a channel in which *p*-aminobenzoic acid binds. In the *M. leprae* there is dapsone resistance and these resistance-causing mutations are located in this channel (Baca *et al.*, 2000). They also speculated that the resistance causing-mutations may prevent dapsone from binding or blocking the channel.

The simulations showed that Pf Ala 437, which is situated at the beginning of loop 2, undergoes

some movement and may assist the substrate (*p*-aminobenzoic acid) with the entry process through hydrophobic interactions. Sulfadoxine may not completely enter the channel and thus needs hydrophobic interaction with Ala 437 to remain in the channel. A mutation would cause the loss of this interaction. The function of the Arg 54 in *M. tuberculosis* may have been taken over by Ala 437 in a reduced capacity (due to the lack of a positive charge on Ala 437).

3.4.1.2. Effects of Water

Baca *et al.* (2000) and Hampele *et al.* (1997) both mention the effect of water in DHPS. Hampele *et al.* (1997) observed that five water molecules occupy the *S. aureus* active site in the unbound state. Four of the water molecules are displaced when the substrate binds. *S. aureus* contains two water molecules in the active site, one which interacts with the Mn^{2+} ion and one which mediates interaction between Asp 167 (Mtb Asp 177, Ec Asp 185, Pf Asp 216, Ba Asp 184) and the pterin moiety. The Asp-water-pterin interaction is preserved in *M. tuberculosis* and *B. anthracis* DHPS. The pyrophosphate groups in the *M. tuberculosis* structure show extensive interactions with four water molecules. The *B. anthracis* structure shows at least one interaction between the pyrophosphate group and a water molecule. In the fully solvated *P. falciparum* model the pyrophosphate group shows polar interactions with at least nine water molecules. The Mg^{2+} ion also lacks an interaction to complete its coordination state but this role may be fulfilled by a water molecule located nearby (as in the *S. aureus* structure, not shown). The presence of conserved water molecules in the active site lends support for implementing protein solvation during molecular dynamics simulations. It also highlights the important function of water in the active site cavity and the necessity to account for water during modelling and *in silico* structural investigations.

3.4.1.3. Explaining Sulfa-drug Resistance in *P. falciparum* DHPS

Korsinczky *et al.* (2004) presented a *P. falciparum* and a *P. vivax* DHPS model but without any insertions or substrates. As sulfadoxine might inhibit DHPS without any substrate present, their model might be useful in analyzing the sulfadoxine-DHPS interaction in a substrate-free environment. The model presented in this study focuses more on the resistance-causing mechanism as related to the DHPS-DHP-sulfadoxine complex. In the model of Korsinczky *et al.* (2004) they predict that Ala 437 is involved in direct hydrophobic contact with sulfadoxine but not *p*-aminobenzoic acid. Although the simulations presented here showed partial movement of loop 2 (the loop on which Ala 437 is located), the loop may move even closer over the active site and thus Ala 437 may move even closer to sulfadoxine.

The A437G mutation may result in loop 2 (now containing Gly 437) becoming more flexible as Gly allows free rotation around the carbon-carbon bond. This increase in flexibility may decrease the interaction of loop 2 (Gly 437) with sulfadoxine and hence a loss of interaction occurs. In the K540E mutation the interaction between the sidechain of residue 540 and the methoxy groups of sulfadoxine

is lost. This results in the hydrophobic cavity being increasingly exposed to solvent and sulfadoxine losing hydrophobic associations with DHPS through residue 540. This is coupled with a charge reversal although the more important change is in the difference in sidechain length between Lys and Glu. The A581G mutation causes a possible loss of hydrophobic interaction between Gly 581 and sulfadoxine (not shown). The S436A mutation may result in the possible loss of a polar interaction between residue 436 and sulfadoxine while the S436F mutation results in sterical clashes as Phe 436 protrudes into the active site (not shown).

From the simulations it seems that although sulfadoxine still binds in the active site, most, if not all of the mutations and mutation combinations, result in a loss of stabilizing interactions between sulfadoxine and DHPS. This would result in sulfadoxine binding less tightly in the active site and thus decreasing the effectiveness of the drug.

3.4.2. *P. falciparum* PPPK

The simulations of the PfPPPK model indicated that the model without the parasite-specific inserts tended to be unstable as the enzyme seemed to "fall apart" after a few tens of picoseconds. This may reveal a stabilizing effect that inserts have on the enzyme as well as a dependency between the enzymes in the bifunctional complex. The simulations also confirmed the large loop movement involved in catalysis in PPPK (Yan *et al.,* 2001). Blaszczyk *et al.* (2000) indicated that the following residues are involved in loop stabilization in *B. anthracis* DHPS: Asn 10 (Pf Asn 25), Gln 50 (Pf Leu 65), Arg 84 (Pf Asn 195), Arg 88 (Pf Glu 200), Trp 89 (Pf Lys 201), Arg 92 (Pf Arg 205) and Tyr 116 (Pf His 324). The differences in loop-anchoring residues between *P. falciparum* and *E. coli* PPPK may have caused the loop interactions to become less stable, and thus the parasite-specific inserts may help to stabilize these loops to ensure proper functioning of the enzyme.

Blaszczyk *et al.* (2000) also refer to several roles for water molecules in the PPPK enzyme. Water molecules may act as a general base to deprotonate the hydroxyl group on HP as well as mediate interactions between PPPK and ATP. Another interesting fact noted by Blaszczyk *et al.* (2000) is the occurrence of several isozymes of PPPK. They attributed this to non-conserved substrate-interacting residues in the PPPK active site. Table 3.1 shows the residues which differ between species.

From Table 3.1 it is evident that the PPPK enzymes of *P. chabaudi, P. yoelli yoelli* and *P. berghei* use the same residues to interact with the substrate (supporting the hypothesis that they might be closely related; Kedziersky *et al.,* 2002) while *P. vivax* seems to use at least three different residues to interact with the substrate. This may indicate that *P. vivax* contains a different isozyme than *P. chabaudi, P. yoelli yoelli* and *P. berghei. P. falciparum* contains a different isozyme as it shares a Glu at position 313 with *P. vivax* but not with the other three Plasmodia. As well as revealing new substrate-interacting residues, this evidence may indicate the existence of different PPPK isozymes in the Plasmodial family.

Table 3.1: The non-conserved substrate-binding residues in PPPK. Blaszczyk *et al.* (2000) attributed the occurrence of PPPK isozymes to active site residue differences between species. Bold typeface indicates inter-plasmodial differences. The one-letter abbreviations of the amino acids are used e.g. 5T means 5 instances of threonine occurring in other species. Adapted from Blaszczyk *et al.* (2000).

Residue in *E.coli* PPPK	**Variation in other species**	**PfPPPK**	**Pv**	**Pb**	**Py**	**Pc**
T42	5T, 3S, 1K, 1N, 1A	T57	T	T	T	T
P43	4P, 2E, 2D, 1K, 1R, 1A	V58	V	V	V	V
L45	2L, 3W, 2F, 1M, 1V, 1Y, 1A	E60	E	E	E	E
Y53	3Y, 8F	F161	F	F	F	F
Q74	5Q, 3L, 2K, 1F	K185	K	K	K	K
R84	8R, 2K, 1D	N195	**E**	N	N	N
R88	6R, 3K, 1H, 1 deletion	E200	E	E	E	E
W89	7W, 1Y, 1D, 2 deletions	K201	**I**	K	K	K
I98	9I, 1V, 1L	I211	I	I	I	I
R110	4R, 3D, 1Q, 1S, 1N, 1I	**E313**	**E**	K	K	K
T112	5T, 3I, 1V, 1A, 1N, 1I	L315	L	L	L	L
Y116	2Y, 7P, 1R, 1E	H319	H	H	H	H
F123	10F, 1S	Y326	Y	Y	Y	Y

3.5. Conclusion

The results obtained for *P. falciparum* PPPK and DHPS highlight the importance of taking loop motion into consideration when studying proteins. In the case of PfDHPS the resistance-causing mutations mostly resided on the loops (loops 2 and 5) and thus their motion had to be studied as well. The investigation of a static model (Chapter 2) would have had difficulty in showing the dynamic range of interactions between PfDHPS, sulfadoxine and DHP. A static model would also have had difficulty in explaining how PfDHPS adapts to the mutations and the effects that the mutations have on the enzyme. The dynamic model revealed the increase in movement of loop 5 with the K540E mutation, as well as the loss of interaction between Glu 540 and sulfadoxine. The dynamic model also revealed how the other mutations affect the binding and stability of sulfadoxine in the active site. The previous two chapters revealed that the mechanism behind sulfadoxine resistance in PfDHPS is due to various factors such as a decrease in interaction between sulfadoxine and DHPS as well as an increase in the motion of the loops.

The simulations of PfPPPK movement highlighted the potential structural importance of the parasite-specific insertions in the *Plasmodium* genome. The simulations showed that a PfPPPK model without the inserts was not stable over long simulations (longer than 100 ps). This points to the stabilizing effect that the inserts and the other bifunctional partner may have on enzyme function and structure.

Chapter 4

Concluding Discussion

Malaria remains one of the deadliest diseases in sub-Saharan Africa. Various drugs have been used as treatment but resistance has been evolving against all of the known drugs. One of the more effective drugs was the pyrimethamine/sulfadoxine combination, which target the DHFR and DHPS enzymes of *Plasmodium falciparum,* respectively. DHPS is part of the folate metabolic pathway and forms a bifunctional enzyme with PPPK. It participates in the production of folate through the addition of a *p*-aminobenzoic acid moiety. Sulfadoxine is a *p*-aminobenzoic acid analogue and thus actively competes for the DHPS active site with *p*-aminobenzoic acid. Although it has been found that amino acid mutations are responsible for resistance of DHPS to sulfadoxine, the mechanism behind it has never been elucidated. One of the main aims of this study was to investigate the mechanism behind sulfadoxine resistance through an *in silico* modelling approach. Additional aims were to propose structures and to investigate the movement of the loops for both bifunctional partners.

As malarial proteins are difficult to crystallize, an alternative approach was followed to obtain 3D structures. Homology modelling was used to obtain models for PfDHPS and PfPPPK. For PfDHPS the crystal structures of *M. tuberculosis* and *B. anthracis* were used as templates as they contained substrate and product analogues in the crystal structures. By comparing the sequences of *P. falciparum, B. anthracis* and *M. tuberculosis* it was evident that *P. falciparum* contained two inserts of about 10 and 32 residues each. It was decided to exclude the longer insert from the modelling process as it did not have any template match. The sequences of *P. berghei, P. chabaudi, P. vivax* and *P. yoelli yoelli* were included to identify *Plasmodium*-specific conserved regions. The crystal structures of *S. aureus* and *E. coli* were available but they were only used for additional protein-substrate interaction information and not as templates (due to low sequence identity and missing loops in the crystal structures). *M. tuberculosis* also showed the highest sequence identity to *P. falciparum* DHPS (30%). Models containing the product analogue as well as sulfadoxine, were built and the product analogue modified to the two natural substrates, *p*-aminobenzoic acid and 6-hydroxymethyldihydropterin pyrophosphate. The protein-substrate complex was subjected to energy minimization to relieve local steric clashes, followed by solvation with TIP3 water molecules. The solvated protein-substrate complex was then minimized again and subsequently

underwent a molecular dynamics simulation for 1ns. The same process was repeated for the construction of the models containing each of the resistance-causing mutations but the molecular dynamics simulation was terminated after 100ps.

For PfPPPK the crystal structure of *E. coli* was used as template. The PfPPPK sequence once again showed insertions when compared against the sequences from other species. In the case of PfPPPK there were two insertions with a length of about 90 residues each. Both of the inserts were excluded from the modelling process. The crystal structure of *E. coli* contained a substrate analogue as well as ATP in the active site, and these were included in the final model. The substrate analogue was modified to the natural substrate, 7,8-dihydro-6-hydroxymethyl-7-methyl-7-pterin (HP), and the protein-substrate complex subjected to energy minimization, followed by solvation and molecular dynamics simulation for 1 ns.

The PfDHPS model was shown to be stable for the duration of the simulation, increasing confidence in the model. During the simulation some of the loops showed movement. This correlates well with the observation that mobile loops cannot be resolved into one consensus orientation in X-ray crystallography, as seen in the crystal structures of DHPS from *M. tuberculosis* (Baca *et al.*, 2000), *S. aureus* (Hampele *et al.*, 1997) and *B. anthracis* (Babaoglu *et al.*, 2004). The simulation of PfDHPS indicated that the substrates remained in the active site for most of the simulation, supporting the accuracy of the active site. The magnesium ion also remained in the active site and anchored 6-hydroxymethyldihydropterin pyrophosphate throughout the simulation. This provides additional support for the accuracy of the active site. Sulfadoxine is a *p*-aminobenzoic acid analogue and thus sulfadoxine and 6-hydroxymethyldihydropterin pyrophosphate were included in the active site to investigate the mechanism of sulfadoxine inhibition and resistance. The simulation indicated that sulfadoxine, due to its larger size, showed more interactions with the protein than *p*-aminobenzoic acid and thus remained in the active site for longer. This suggested possible competition between sulfadoxine and *p*-aminobenzoic acid for the active site in the wild-type protein. Sulfadoxine showed interactions with Lys 540, Ala 581, Ser 436 and Ala 437. The position of *p*-aminobenzoic acid in the DHPS active site has not been resolved and experimental work such as the mutation of Pf Lys 609 (proposed to interact with *p*-aminobenzoic acid; Babaoglu *et al.*, 2004) may prove to be valuable. Derremaux *et al.* (1998) suggested that loop movement might be the determining factor in the time taken for catalysis to occur in TIM-barrel proteins (such as DHPS). Decreasing the mobility of the loops in PfDHPS through mutation studies should shed some light on the effect of the loops on catalysis and enzyme kinetics.

During the simulations of the different mutants, it was shown that each resistance-causing mutation (Ser436Ala, Ala436Gly, Lys540Glu and Ala581Gly) decreased the interaction between sulfadoxine and PfDHPS. When multiple mutations occurred the effect was additive. The simulation showed that *p*-aminobenzoic acid did not have major interactions with these residues and thus the normal functioning

Table 4.1: The similarities between pterin-active site interactions in PfDHPS and PfPPPK. Note that the ligand's hydrogen bond-forming capacity is saturated in both DHPS and PPPK (Figure 2.6 and 2.14).

Ligand Atoms	**Type of interaction**	**PfDHPS**	**PfPPPK**
All C atoms	Hydrophobic interactions	Phe 221, Ile 145, Met 170	Phe 68, Tyr 141
All N atoms	Hydrogen bonds to mostly charged residues as well as residues containing oxygen groups in the side-chain	Asp 123, Asn 143, Asp 216	Glu 60 (side chain as well as backbone atoms), Thr 57, Val 58 (backbone)
O atom	Hydrogen bond with amino groups	Lys 250	Asn 70

of the enzyme could continue. Using the model as a guide, mutations which prevent *p*-aminobenzoic acid from entering the active site may prove valuable in further validating the model.

The dynamics simulation of PfPPPK was less revealing than that of PfDHPS but some valuable information was still obtained. ATP and HP substrate actually moved closer to one another during the simulation as would be expected during normal catalysis. Support for the predicted active site conformation was gained from the fact that the two magnesium ions remained in the active site and anchored the ATP throughout the simulation. Yan *et al.* (2001) showed that HiPPPK has extensive loop movement, as seen from the different conformations of the protein with and without ligands in the active site. Their work also indicated that PPPK might work through an induced fit model. PfPPPK showed extensive loop movement during the initial part of the simulation and this supports the work done by Yan *et al.* (2001) on the mobility of the loops. The work presented here also supports the theory of induced fit in PPPK as proposed by Yan *et al.* (2001) for HiPPPK. The only drawback of the simulation was that the model began to loose tertiary structure towards the end of the simulation (close to 0.7ns). This may be due to a stabilizing effect the parasite-specific insertions have on the enzyme as well as the fact that PPPK is missing its bifunctional partner, DHPS. Insert stabilization could be shown by modelling the complete bifunctional enzyme but since a template for the linking region and the inserts are not available, this approach was not pursued further. If the *P. falciparum* PPPK-DHPS crystal structure would become available, molecular dynamics of the bifunctional complex with and without the inserts would prove invaluable in determining the effect of the inserts on enzyme stabilization.

The occurrence of multiple drug resistance in *P. falciparum* necessitates the need for new drugs as well as an understanding of the mechanism behind current drug resistance in malaria. The studies reported here have contributed to the understanding of the mechanism behind sulfadoxine resistance in *P. falciparum* DHPS. Both the models have been shown to be stable (DHPS more so than PPPK) and could thus be used in investigating the active sites with the aim of designing new drugs, which can target DHPS more effectively. Another approach is to make a single drug which targets more than one enzyme. PfDHPS and PfPPPK both bind substrates with a similar backbone structure (pterin group)

in a similar way and thus the models presented here would help in revealing similarities between the two active sites (Table 4.1). These similarities could be exploited to design a pterin backbone-based drug which targets both enzymes at the same time. A drug which resembles the product, pteroic acid, would also prove valuable as it would also target DHPS specifically. The work presented here also touches on the importance of investigating non-static models when looking at drug resistance. A static model would not have revealed the subtle effects each mutation has on sulfadoxine binding in PfDHPS.

Overall the work presented here provides models for *P. falciparum* PPPK and DHPS as well as representations of their respective active sites. This was used to propose a mechanism for sulfadoxine resistance, which would have been difficult to elucidate only with experimental investigations. The study also supported work by Yan *et al.* (2001) and Derremaux *et al.* (1998) who showed that loop movements are important for shielding the substrate from the solvent in PPPK and DHPS, respectively, as well as playing a role in enzyme catalytic rate.

Possible further work on this enzyme system would include molecular dynamics studies of the effects of the inserts on enzyme stabilization (when the crystal structure becomes available) and mutation studies to investigate the effect of loop movement in DHPS on enzyme catalytic rates. Other structural work such as NMR to investigate at the possible coordination of loop movement in the PPPK and DHPS complex as well as secondary structure in the inserts would prove valuable. Rational drug design could be used to design inhibitors against both enzymes and deletion studies would help in clarifying the function of the parasite-specific inserts. The work presented here should help in understanding the mechanism behind sulfadoxine resistance in *P. falciparum* DHPS and help to identify new directions in inhibitor research against malaria.

Summary

Malaria kills nearly 1.5 million and affects more than 500 million people annually, mostly in sub-Saharan Africa. The malaria parasite has developed resistance against almost all of the known drugs used for treatment. This fact has resulted in a constant battle between developing new anti-malarials and the parasite evolving resistance. One of the main drug combinations, pyrimethamine/sulfadoxine, targets the dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS) proteins in the folate synthesis pathway of human malaria parasite, *Plasmodium falciparum*. The folate synthesis pathway is absent from the human host and thus presents itself as an ideal target for parasite-specific drugs.

The three dimensional atomic coordinates of a target protein can help in designing new, more effective drugs. Malarial proteins are notoriously difficult to crystallize and thus homology modelling was chosen as an alternative method to obtain a protein structure. DHPS and PPPK occur as a bifunctional protein in the folate metabolism pathway. In this study, homology modelling was used to do *in silico* modelling of *P. falciparum* DHPS and hydroxymethyldihydropteridine pyrophosphokinase (PPPK). For the *P. falciparum* DHPS model the crystal structures of *M. tuberculosis* and *B. anthracis* DHPS were used as templates and for the *P. falciparum* PPPK model, the crystal structure of *E. coli* PPPK. Molecular dynamics was used to investigate loop movement in DHPS and PPPK as well as to reveal the effect of resistance-causing mutations on sulfadoxine binding in *P. falciparum* DHPS.

This study revealed that four of the five known sulfadoxine resistance-causing mutations in DHPS disrupt the interaction between sulfadoxine and DHPS. This translates to a reduced capacity for sulfadoxine to inhibit DHPS, and results in resistance. The simulations also showed that both DHPS and PPPK have extensive loop movements during catalysis. The loop movements in DHPS and PPPK may also play a role in determining the catalytic rate of the enzymes.

The work presented here provides researchers with models of *P. falciparum* DHPS and PPPK. These models can be used to design experiments to investigate resistance, design new drugs and probe the structure of the PPPK-DHPS bifunctional enzyme.

Bibliography

Achari, A., Somers, D. O., Champness, J. N., Bryant, P.K., Rosemond, J. and Stammers, D. K. (1997) Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nature Structural Biology*, **4**, 490-497.

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **3**, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST adn PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25**, 3389-4302.

Aravind, L., Iyerm, L. M., Wellems, T. E. and Miller, L. H. (2003) *Plasmodium* Biology: Genomic gleanings. *Cell,* **115**, 771-785.

Babaoglu, K., Qi, J., Lee, R. E. and White, S. W. (2004) Crystal Structure of 7,8-Dihydropteroate Synthase from *Bacillus anthracis*: Mechanism and Novel Inhibitor Design. *Structure,* **12**, 1705-1717.

Baca, A. M., Sirawaraporn, R., Turley, S., Sirawaraporn, W. and Hol W. G. J. (2003) Crystal structure of *Mycobacterium tuberculosis* 6-hydroxymethyl-7,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action. *Journal of Molecular Biology*, **302**, 1193-1212.

Bailey, T. L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, **2**, 28-36.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. and Eddy SR. (2004) The Pfam protein

families database. *Nucleic Acids Research,* **32**, D138-D141.

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. *Biochemical Journal*, **377**, 439-448.

Blaszczyk, J., Shi, G., yan, H. and Ji, X. (2000) Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 Å resolution. *Structure*, **8**, 1049-1058.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement, TrEMBL in 2003. *Nucleic Acids Research,* **31**, 365-370.

Bond, C. S. (2003) Topdraw: a sketchpad for protein structure topology cartoons. *Bioinformatics,* **19**, 311-312.

Bonnet, J., Corbel, V., Darriet, F., Chandre, F. and Hougard, J .M. (2004) Topical applications of pyrethroid and organophosphate mixtures revealed positive interactions against pyrethroid-resistant *Anopheles gambiae*. *Journal of the American Mosquito Control Association,* **20**, 438-443.

Brabin, L. and Brabin, B. J. (2005) HIV, malaria and beyond: reducing the disease burden of female adolescents. *Malaria Journal*, **4**, 2.

Brink, J., Ludtke, S. J., Kong, Y., Wakil, S. J., Ma, J. and Chiu, W. (2004) Experimental Verification of Conformational Variation of Human Fatty Acid Synthase as Predicted by Normal Mode Analysis. *Structure,* **12**, 185 191.

Brower, R. C., Vasmatzis, G., Silverman, M. and DeLisi, C. (1993) Exhaustive conformational search and simulated annealing for models of lattice peptides. *Biopolymers*, **33**, 329-334.

Bruccoleri, B. R. (1993) Application of systematic conformation search to protein modelling. *Molecular Simulation*, **10**, 151-174.

Bruccoleri, B. R. and Karplus, M. (1990) Conformational sampling using high temperature molecular

dynamics. *Biopolymers*, **29**, 1847-1862.

Bwijo, B., Kaneko, A., Takechi, M., Zungu, I. L., Moriyama, Y., Lum, J. K., Tsukahara, T., Mita, T., Takahashi, N., Bergqvist, Y., Björkman, A. and Kobayakawa, T. (2003) High prevalence of quintuple mutant *dhfr/dhps* genes in *Plasmodium falciparum* infections seven years after introduction of sulfadoxine and pyrimethamine as first line treatment in Malawi. *Acta Tropica*, **85**, 363-373.

Bystroff, C., Thorsson, V. and Baker, D. J. (2000) HMMSTR: a hidden markov model for local sequence-structure correlations in proteins. *Journal of Molecular Biology,* **301**, 173-190.

Charlwood, J. D., Pinto, J., Ferrara, P. R., Sousa, C. A., Ferreira, C., Gil, V. and do Rosario, V. E. (2003) Raised houses reduce mosquito bites. *Malaria Journal,* **2**, 45-51.

Chothia, C. and Lesk, A. M. (1987) Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology*, **196**, 901-917.

Copley, R. R. and Bork, P. (2000) Homology among $(\beta\alpha)_8$ barrels: implications for the evolution of metabolic pathways. *Journal of Molecular Biology,* **303**, 627-640.

Cruz Marques, A. (1987) Human migration and the spread of malaria in Brazil. *Parasitology Today,* **3**, 166-170.

De Groot, B. L. and Grubmüller, H. (2001) Water permeation across biological membranes: Mechanism and dynamics of Aquaporin-1 and GlpF. *Science,* **294**, 2353-2356.

DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

Derremaux, P. and Schlick, T. (1998) The loop opening/closing motion of the enzyme triosephosphate isomerase. *Biophysical Journal*, **74**, 72-81.

Dougall, D. S., Gopalakrishnan, V., Hennessey, D. and Rosenberg, J. (2004) Mining the PDB for information leading to successful protein crystallization experimental design. ISMB 2004 conference, Glasgow, poster presentation.

Eckstein-Ludwig, U., Webb, R. J., van Goethem, I. D. A., East, J. M., Lee, A. G., Kimura, M., O'Neill, P. M., Bray, P. G., Ward, S. A. and Krishna, S. (2003) Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature,* **424**, 957-961.

Evans, J. S., Mathiowetz, A. M., Chan, S. I. and Goddard III, W. A. (1995) De novo prediction of polypeptide conformations using dihedral probability grid Monte Carlo methodology. *Protein Science,* **4**, 1203-1216.

Ewald, P. (1921) Due berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen den Physik,* **64**, 253-287.

Fersht, A. R. (2002) On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *Proceedings of the National Academy of Sciences of the United States of America,* **99**, 14122-14125.

Ferreira, M. U., Ribeiro, W. L., Tonon, A .P., Kawamoto, F. and Rich, S. M. (2003) Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*. *Gene,* **304**, 65-75.

Fiser, A., Do, R. K. and Sali A. (2000) Modeling of loops in protein structures. *Protein Science,* **9**, 1753-1773.

Friedman, H. L. (1975) Image approximation to the reaction field. *Molecular Physics*, **29**, 1533-1543.

Friedman, J. F., Philips-Howard, P. A., Hawley, W. A., Terlouw, D. J., Kolczak, M. S., Barber, M., Okello, N., Vulule, J. M., Duggan, C., Nahlen, B. L. and Ter Kuile, F. O. (2003) Impact of permethrin-treated bed nets on growth, nutritional status and body composition of primary school children in western Kenya. *American Journal of Tropical Medicine and Hygiene,* **68**, 78-85.

Genton, B., Betuela, I., Felger, I., Al-Yaman, F., Anders, R. F., Saul, A., Rare, L., Baisor, M., Lorry, K., Brown, G. V., Pye, D., Irving, D. O., Smith, T. A., Beck, H-P. and Alpers, M. P. (2002) A Recombinant Blood-Stage Malaria Vaccine Reduces *Plasmodium falciparum* Density and Exerts Selective Pressure on Parasite Populations in a Phase 1-2b Trial in Papua New Guinea. *The Journal of Infectious Diseases,* **185**, 820-827.

Genton, B., Al-Yamana, F., Betuelaa, I., Andersc, R. F., Saule, A., Baeaa, K., Mellomboa, M., Taraikaa, J., Brown, G. V., Pyef, D., Irving, D. O., Felgerb, I., Beck, H-P., Smith, T .A. and Alpersa, M. P. (2003) Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against *Plasmodium falciparum* in Papua New Guinean children. *Vaccine,* **22**, 30-41.

Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics,* **19**, 1015-1018.

Gilman, A. G., Hardman, J. G. and Limbird, L. E. (2001) In: *Goodman & Gilman's The pharmacological basis of therapeutics.* Tenth edition, McGraw-Hill, New York.

Greer, J. (1980) Model for haptoglobin heavy chain based upon structural homology. *Proceedings of the National Academy of Sciences of the United States of America,* **Homology modelling,** 3393-3397.

Grundy, W. N., Bailey, T. L., Elkan, C .P. and Baker, M .E. (1997) Meta-MEME: motif-based hidden Markov models of protein families. *Computer Applications in the Biosciences,* **13**, 397-406.

Hampele, I. C., D'Arcy, A., Dale, G. E., Kostrewa, D., Nielsen, J., Oefner, C., Page, M. G., Schonfeld, H. J., Stuber, D. and Then, R. L. (1997) Structure and function of the dihydropteroate synthase from *Staphylococcus aureus. Journal of Molecular Biology,* **268**, 21-31.

Hartl, D. L. (2004) The origin of malaria: mixed messages from genetic diversity. *Nature Reviews,* **2**, 15-22.

Henning, M., Dale, G. E., D'Arcy, A., Danel, F., Fischer, S., Gray, C. P., Jolidon, S., Muller, F., Page, M. G., Pattison, P. and Oefner, C. (1999) The structure and function of the 6-hydroxymethyl-7,8-dihydropterin pyrophosphoskinase from *Haemophilus influenzae. Journal of Molecular Biology,* **26**, 211-219.

Hooft, R. W. W., Sander, C. and Vriend, G. (1996) Verification of protein structures: sidechain planarity. *Journal of Applied Crystallography*, **29**, 714-716.

Hubbard, T. J. P., Ailey, B., Brenenr, S. E., Murzin, A. G. and Chothia, C. (1999) SCOP: A structural classification of proteins database. *Nucleic Acids Research,* **27**, 254-256.

Hughes, A. L. and Verra, F. (2001) Very large long-term effective population size in the virulent

malaria parasite *Plasmodium falciparum. Proceedings of the Royal Society of London. Series B. Biological sciences,* **268**, 1855-1860.

Inselburg, J. and Zhang, R. D. (1988) Study of dihydrofolate reductase-thymidylate synthase in *Plasmodium falciparum. American Journal of Tropical Medicine and Hygiene,* **39**, 328-336.

IUBMB (1992)In: *Enzyme Nomenclature.* Academic Press, San Diego.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. and Lecomte, C. (2000) Accurate protein crystallography at ultra-high resolution: Valence electron distribution in crambin. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 3171-3176.

Joët, T. and Krishna, S. (2004) The hexose transporter of *Plasmodium falciparum* is a worthy drug target. *Acta Tropica,* **89**, 371-374.

Jones, G.,Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology,* **267**, 727-748.

Joy, D. A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A. U., Ho, M., Wang, A., White, N. J., Suh, E., Beerli, P. and Su, X. Z. (2003) Early origin and recent expansion of *Plasmodium falciparum. Science,* **300**, 318-321.

Kantardjieff, K. A. and Rupp, B. (2004) Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics*, **20**, 2162-2168.

Kasekarn, W., Sirawaraporn, R., Chahomchen, T., Cowman, A. F. and Sirawaraporn, W. (2004) Molecular characterization of bifunctional hydroxymethyldihydropterin pyrophosphokinasedihydropteroate synthase from *Plasmodium falciparum. Molecular and Biochemical Parasitology,* **137**, 43-53.

Kedzierski, L., Escalante, A. A., Isea, R., Black, C. G., Barnwell, J. W. and Coppel, R. L. (2002) Phylogenetic analysis of the genus *Plasmodium* based on the gene encoding adenylosuccinate lyase. *Infection, Genetics and Evolution*, **1**, 297-301.

Kilama, W. L. (2003) Malaria vaccines in Africa. *Acta Tropica,* **88**, 153-159.

King, R. D. and Sternberg, M. J. (1996) Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Science,* **5** (11), 2298-2310.

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. *Antimicrobial Agents and Chemotherapy,* **48**(6), 2214-2222.

Kremsner, P. G. and Krishna, S. (2004) Antimalarial combinations. *The Lancet,* **364**, 285-294.

Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science,* **289**, 1546-1550.

Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* **26**, 283-291.

Lazaridis, T. and Karplus, M. (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *Journal of Molecular Biology,* **288**, 477-487.

Leach, A. R. (2001) In: *Molecular Modelling: Principles and Applications.* Second edition, Prentice Hall.

Levitt, M. (1992) Accurate modeling of protein conformation by automatic segment matching. *Journal of Molecular Biology,* **226**, 507-533.

Ma, J., Sigler, P. B., Xu, Z. and Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL. *Journal of Molecular Biology* **302**, 303-313.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure,* **29**, 291-325.

Masimirembwa, C. M., Phuong-dung, N., Phuc, B. Q., Duc-Dao, L., Sy, N. D., Skold, O. and Swedberg, G. (1999) Molecular epidemiology of *Plasmodium falciparum* antifolate resistance in Vietnam: genotyping for resistance variants of dihydropteroate synthase and dihydrofolate reductase. *Interna-*

tional Journal of Antimicrobial Agents, **12**, 203-211.

Mawili-Mboumba, D-P., Ekala, M-T., Lekoulou, F. and Ntoumi, F. (2001) Molecular analysis of DHFR and DHPS genes in *P. falciparum* clinical isolates from the Haute-Ogooue region in Gabon. *Acta Tropica,* **78**, 231-240.

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics,* **16**, 404-405.

MacKerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, III, W., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, **102**, 3586-3616.

McGarrah, D. B. and Judson, R. S. (1993) Analysis of genetic algorithm method for molecular conformation determination. *Journal of Computational Chemistry*, **14**, 1385-1395.

McMurray, J. (2000) In: *Organic chemistry.* Fifth edition, Brooks/Cole.

Mitamura, T. and Palacpac, N. M. Q. (2003) Lipid metabolism in *Plasmodium falciparum*-infected erythrocytes: possible new targets for malaria chemotherapy. *Microbes and Infection,* **5**, 545-552.

Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *Journal of Computational Chemistry*, **19**, 1639-1662.

Mutabingwa, T., Nzila, A., Mberu, E., Nduati, E., Winstanley, P., Hills, E. and Watkins, W. (2001) Chlorproguanil-dapsone for treatment of drug-resistant falciparum malaria in Tanzania. *The Lancet,* **358**, 1218-1223.

Nagano, N., Orengo, C. A. and Thornton, J. M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.

Journal of Molecular Biology, **321**, 741-765.

Neuvirth, H., Raz, R. and Schreiber, G. (2004) ProMate: A structure based prediction program to identify the location of protein-protein binding sites. *Journal of Molecular Biology,* **338**, 181-199.

Nirmalan, N., Sims, P. F. G. and Hyde, J. E. (2004) Translational up-regulation of antifolate drug targets in the human malaria parasite *Plasmodium falciparum* upon challenge with inhibitors. *Molecular and Biochemical Parasitology,* **136**, 63-70.

Nzila, A. M., Mberu, E. K., Sulo, J., Dayo, H., Winstanley, P. A., Sibley, C. H. and Watkins, W. M. (2000) Towards an understanding of the mechanism of pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: genotyping of dihydrofolate reductase and dihydropteroate synthase of Kenyan parasites. *Antimicrobial Agents and Chemotherapy,* **44**, 991-996.

Olliaro, P. (2001) Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacology & Therapeutics,* **89**, 207-219.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. and Thornton, J. M. (1997) CATH - A Hierarchic Classification of Protein Domain Structures. *Structure*, **5**, 1093-1108.

Pinto, A. Y., Azevedo, C. H., da Silva, J. B. and de Souza, J. M. (2003) Assessment of chloroquine single dose treatment of malaria due to *Plasmodium vivax* in Brazilian Amazon. *Revista do Instituto de Medicina Tropical de Sao Paulo,* **45**, 327-331.

Poirot, O., Suhre, K., Abergel, C., O'Toole, E. and Notredame, C. (2004) 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. *Nucleic Acids Research*, **32**, W37-40.

Ranson, H., Paton, M. G., Jensen, B., McCarroll, L., Vaughan, A., Hogan, J. R., Hemingway, J. and Collins, F. H. (2004) Genetic mapping of genes conferring permethrin resistance in the malaria vector, *Anopheles gambiae. Insect Molecular Biology*, **13**, 379-386.

Rich, S. M., Licht, M. C., Hudson, R. R. and Ayala, F. J. (1998) Malaria's eve: evidence of a recent population bottleneck throughout the world population of *Plasmodium falciparum. Proceedings of the*

National Academy of Sciences of the United States of America, **95**, 4425-4430.

Ring, C. S. and Cohen, F. E. (1994) Conformational sampling of loop structures using genetic algorithms. *Israel Journal of Chemistry,* **34**, 245-252.

Roper, C., Pearce, R., Bredenkamp, B., Gumede, J., Drakeley, C., Mosha, F., Chandramohan, D. and Sharp, B. (2003) Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *The Lancet,* **361**, 1174-1181.

Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Engineering,* **12**, 85-94.

Rost, B. and Lui, J. (2003) The PredictProtein server. *Nucleic Acids Research*, **31** (13), 3300-3304.

Rozovsky, S. Jogl, G., Tong, L. and McDermott, A. E. (2001) Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site dynamics. *Journal of Molecular Biology*, **310**, 271-280.

Sakihama, N., Matsuo, T., Mitamura, T., Horii, T., Kimura, M., Kawabata, M. and Tanabe, K. (2004) Relative frequencies of polymorphisms of variation in Block-2 repeats and 5' recombinant types of *Plasmodium falciparum msp1* alleles. *Parasitology International*, **53**, 59-67.

Salcedo, E., Cortese, J. F., Plowe, C. V., Sims, P. F. and Hyde, J. E. (2001) A bifunctional dihydrofolate synthetase–folylpolyglutamate synthetase in Plasmodium falciparum identified by functional complementation in yeast and bacteria. *Molecular and Biochemical Parasitology,* **112**, 239-252.

Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, **234**, 779-815.

Samba, E. M. (2004) Bridging the gap: Linking research, training and service delivery to reduce the malaria burden in Africa. *American Journal of Tropical Medicine and Hygiene,* **71** (supplement), 0.

Schlick, T. (2002) In: *Molecular Modeling and Simulation. An interdisciplinary guide.* First edition, Springer New York.

Sherman, I. W. (1998) In: *Malaria:Parasite biology, pathogenesis and protection.* First edition, ASM

press, Washington.

Snow, R. W., Molyneux, C. S., Njeru, E. K., Omumbo, J., Nevill, C. G., Muniu, E. and Marsh, K. (1997) The effects of malaria control on nutritional status in infancy. *Acta Tropica,* **65**, 1-10.

Stammers, D. K., Achari, A., Somers, D. O'N., Bryant, P. K., Rosemond, J., Scott, D. L. and Champness, J. N. (1999) 2.0 Å X-ray structure of the ternary complex of 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase from *Escherichia coli* with ATP and a substrate analogue. *FEBS Letters*, **456**, 49-53.

Sutcliffe, M. J., Haneef, I., Carney, D. and Blundell, T. L. (1987) Knowledge based modelling of homologous proteins, Part I: Three dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Engineering,* **1**, 377-384.

Tama, F., Miyashita, O. and Brooks, C. L. (2004) Experimental Verification of Conformational Variation of Human Fatty Acid Synthase as Predicted by Normal Mode Analysis. *Journal of Molecular Biology,* **337**, 985-999.

Tang, P and Xu, Y. (2002) Large-scale molecular dynamics simulations of general anesthetic effects on the ion channel in the fully hydrated membrane: the implication of molecular mechanisms of general anesthesia. *Proceedings of the National Academy of Sciences of the United States of America,* **99**, 16035-16040.

Thanki, N., Zeelen, J. P., Mathieu, M., Jaenicke, R., Abagyan, R. A., Wierenga, R. K. and Schliebs, W. (1997) Protein engineering with monomeric triosephosphate isomerase (monoTIM): the modelling and structure verification of a seven-residue loop. *Protein Engineering*, **10**, 159-167.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressively multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

Toxvaerd, S. (1990) Molecular dynamics calculation of the equation of state of alkanes. *Journal of Chemical Physics*, **93**, 4290-4295.

Traut, T. and Temple, B. R. S. (2000) The Chemistry of the Reaction Determines the Invariant Amino Acids during the Evolution and Divergence of Orotidine 5'-Monophosphate Decarboxylase. *The*

Journal of Biological Chemistry, **275**, 28675-28681.

Triglia, T. and Cowman, A. F. (1994) Primary structure and expression of the dihydropteroate synthase gene of *Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America,* **91**, 7149-7153.

Triglia, T., Menting, J. G., Wilson, C. and Cowman, A. F. (1997) Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America,* **94**, 13944-13949.

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L. E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: Chemical shift assignments and secondary structure of malate synthase G. *Journal of the American Chemical Society,* **124**, 10025-10035.

Vaidya, A. B. (2004) Mitochondrial and Plastid functions as antimalarial drug targets. *Current Drug Targets - Infectious Disorders,* **4**, 11-23.

Vasconcelos, K. F., Plowe, C. V., Fontes, C. J., Kyle, D., Wirth, D. F., Pereira da Silva, L. H. and Zalis, M. G. (2000) Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase of isolates from the Amazon region of Brazil. *Memorias do Instituto Oswaldo Cruz,* **95**, 721-728.

Vinnicombe, H. G. and Derrick, J. P. (1999) Dihydropteroate synthase from *Streptococcus pneumoniae*: characterization of substrate binding order and sulfonamide inhibition. *Biochemical and Biophysical Research Communications*, **258** (3), 752-757.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering* **8**, 127-134.

Wang , P., Chung-Shinn, L., Bayoumi, R., Djimde, A., Doumbo, O., Swedberg, G., Duc Dao, L., Mshinda, H., Tanner, M., Watkins, W. M., Sims, F. P. G. and Hyde, J. E. (1997) Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Molecular and Biochemical Parasitology,* **89**, 161-177.

Wang, P., Nirmalan, N., Wang, Q., Sims, P. F. and Hyde, J. E. (2004) Genetic and metabolic analy-

sis of folate salvage in the human malaria parasite *Plasmodium falciparum. Molecular and Biochemical Parasitology,* **135**, 77-87.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *Journal of the American Chemical Society,* **106**, 765-784.

Wierngа, R. K. (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS letters,* **492**, 193-198.

Wongsrichanalai, C., Pickard, A. L., Wernsdofer, W. H. and Meshnick, S. R. (2002) Epidemiology of drug-resistant malaria. *The Lancet,* **2**, 209-218.

Wootton, J. C. and Federhen, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Computers and Chemistry,* **17**, 149-163.

Yan, H., Blaszczyk, J., Xiao, B., Shi, G. and Ji, X. (2001) Structure and dynamics of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Journal of Molecular Graphics and Modelling,* **19**, 70-77.

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanakul, J., Sirawarporn, W., Taylor, P., Walkinshaw, M. D. and Yuthavong, Y. (2003) Insights into antifolate resistance from malarial DHFR-TS structures. *Nature Structural Biology,* **5**, 345-362.