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.

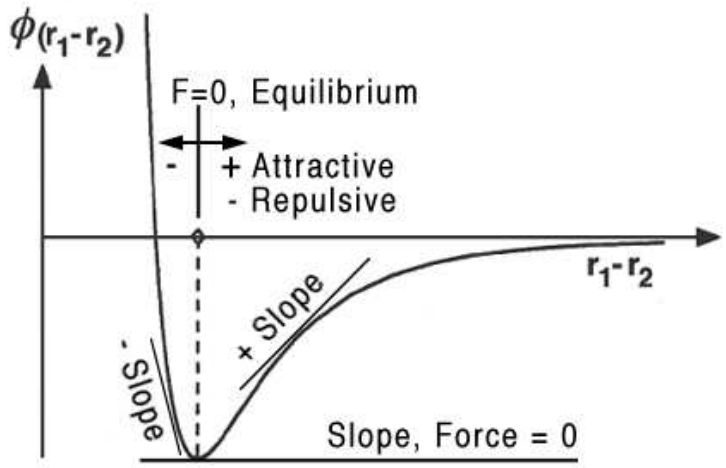

Figure 3.1: The Lennard-Jones potential. This function describes the interaction of two molecules or atoms over a distance. When the attractive and repulsive forces are equal a bond has formed or a stable interaction has occurred. As molecules move further away (r_1-r_2, indicated on the y-axis) from 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.

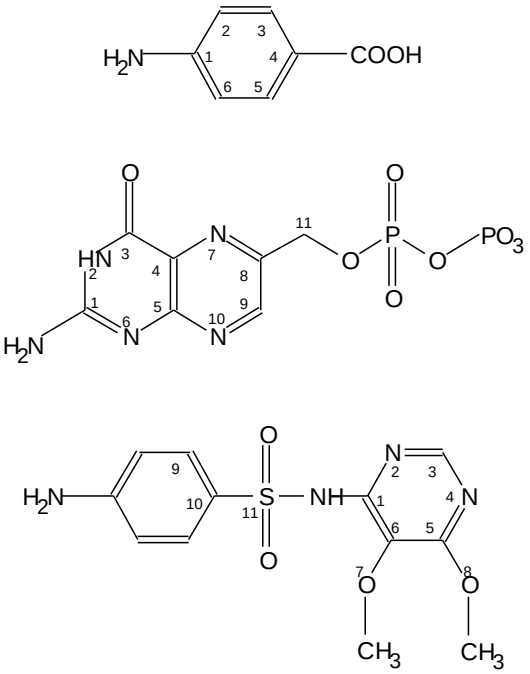

Figure 3.2: The structures of *p*-aminobenzoic acid (top), DHP (middle) and sulfadoxine (bottom). The three ligands were parameterized using closely related atom types and bonds from the the CHARMM parameter files. The dihedral angle energies were then adjusted to keep the appropriate bonds planar. Only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

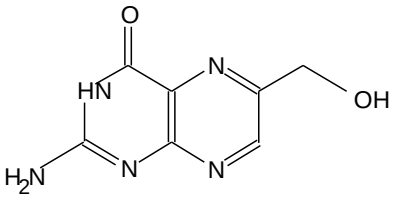

Figure 3.3: The substrate of PfPPPK, 6-hydroxymethyl-7,8-dihydropterin. The OH group is replaced by a pyrophosphate group from the ATP to form DHP (Blaszczyk *et al.*, 2000).

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.

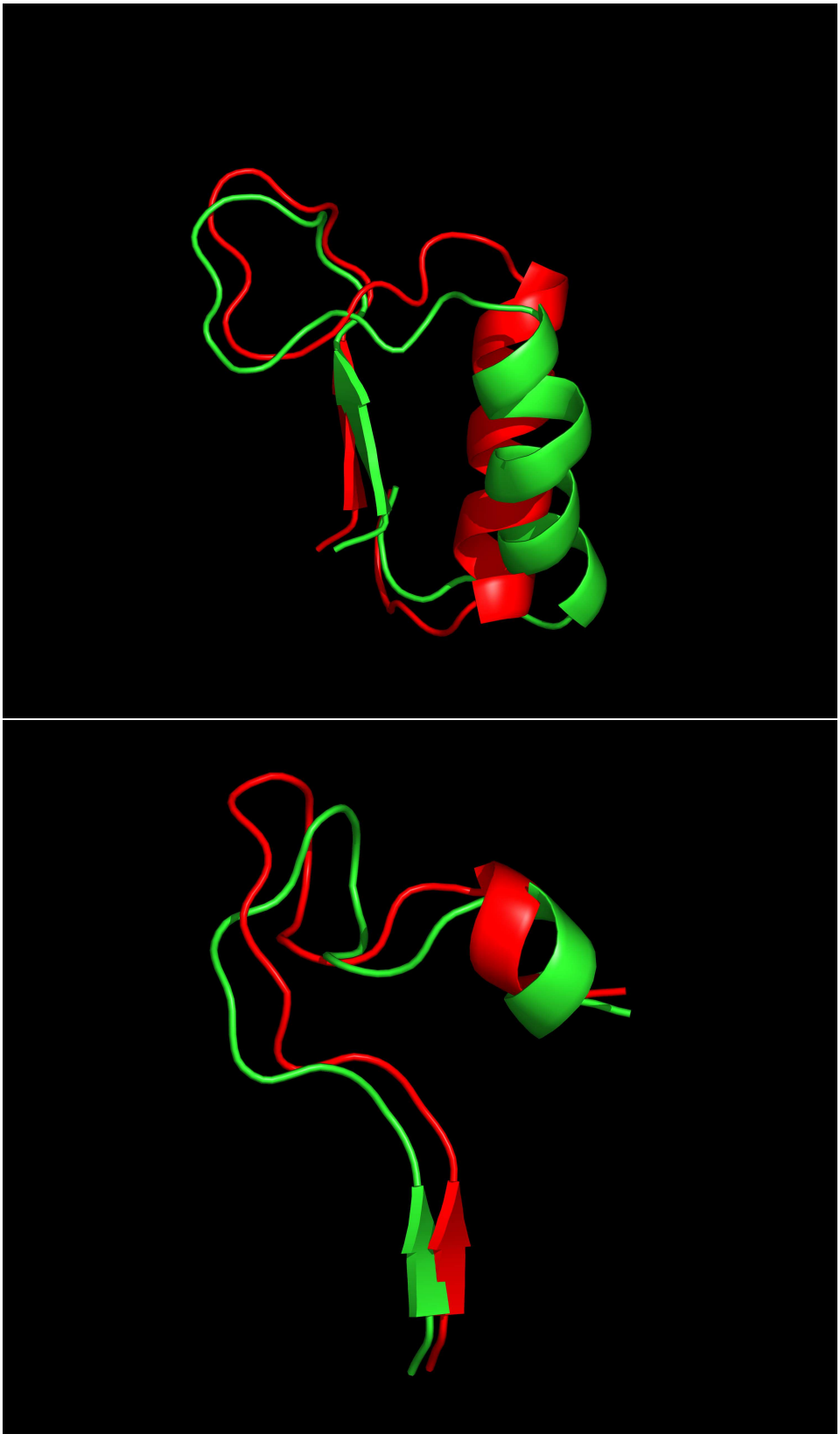

Figure 3.4: Refer to Figure 3.9 for location of various loops. Top: The movement of loop 1 in the PFDHPS model. Bottom: The movement of loop 2 during the simulation. Loop 2 anchors Mg^{2+} and although there is not much movement of the loop during the simulation, the helices move substantially while 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).

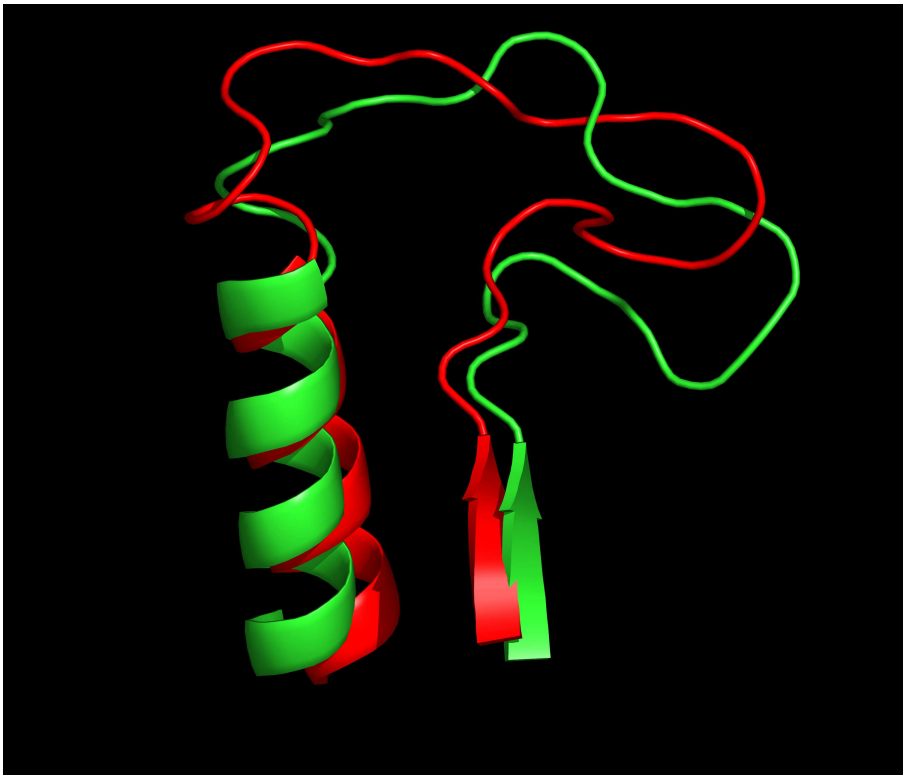

Figure 3.5: The range of movement of loop 5. Note the stability of the secondary structural elements. Generated with PYMOL (DeLano, 2002).

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.

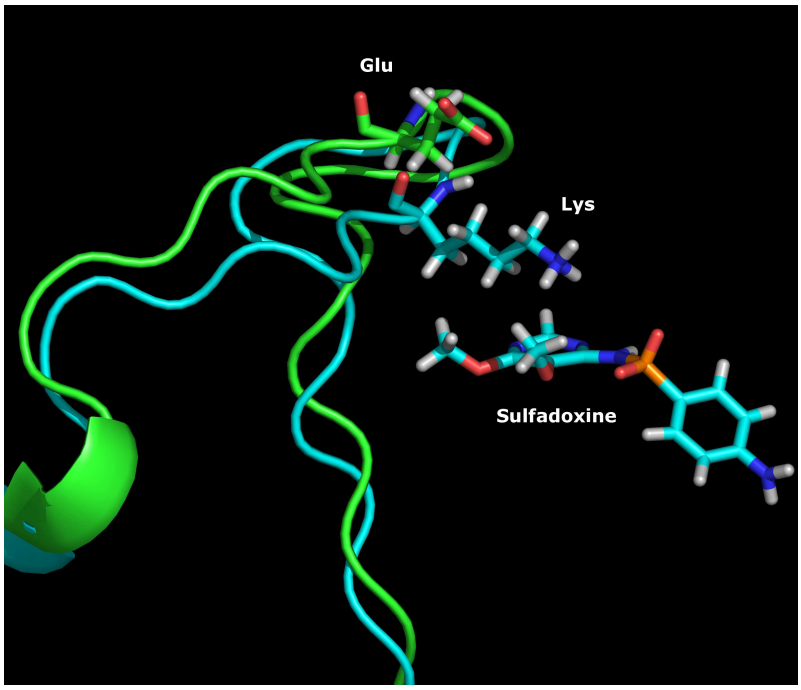

Figure 3.7: The effect of the K540E mutation on sulfadoxine binding. In the wild-type (light blue) the methoxy groups on sulfadoxine show interaction with the sidechain of Lys 540. Although the side chain is protonated there is no interaction between it and the sulfone group on sulfadoxine. When the 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

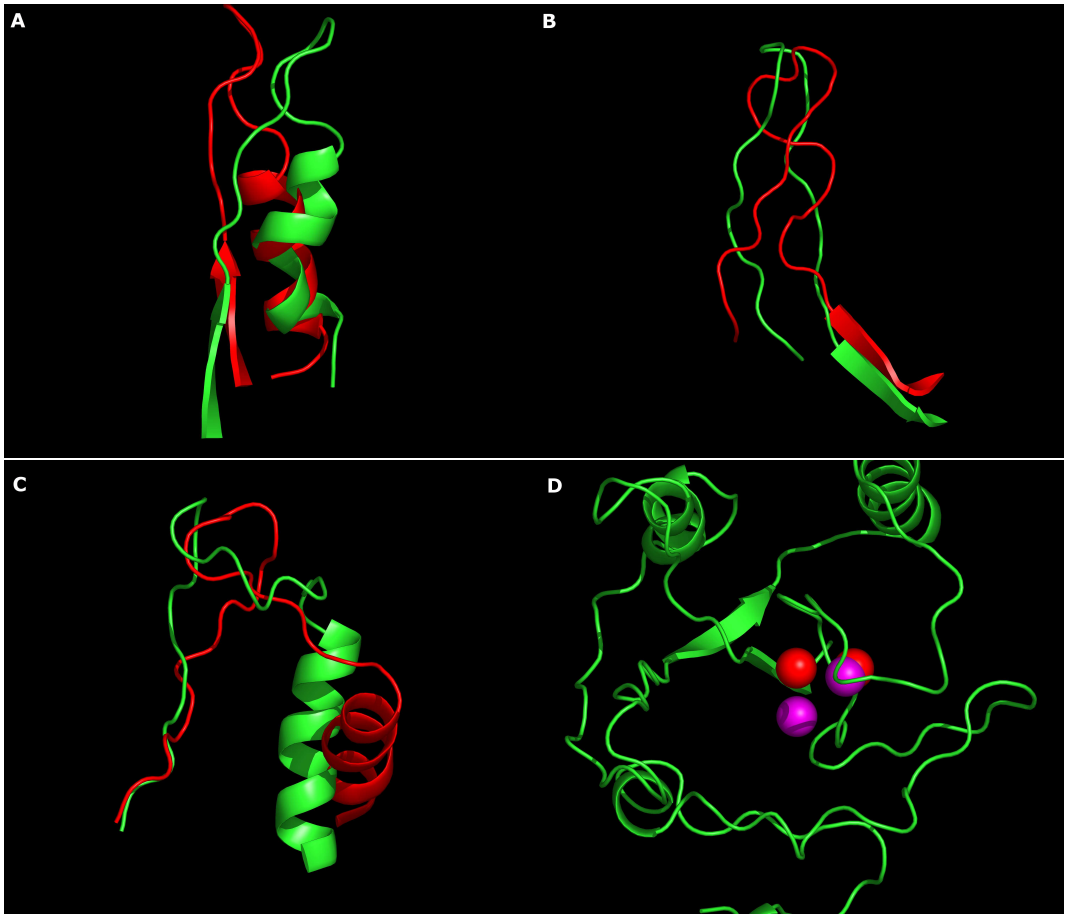

Figure 3.8: The movement of the different loops and the Mg^{2+} ions in the PfPPPK model. The starting position of the protein is indicated in green with red showing the position after the simulation. A: the movement of loop 1. Note the deformation of α-helix 1. B: The movement of loop 2. Once again 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.

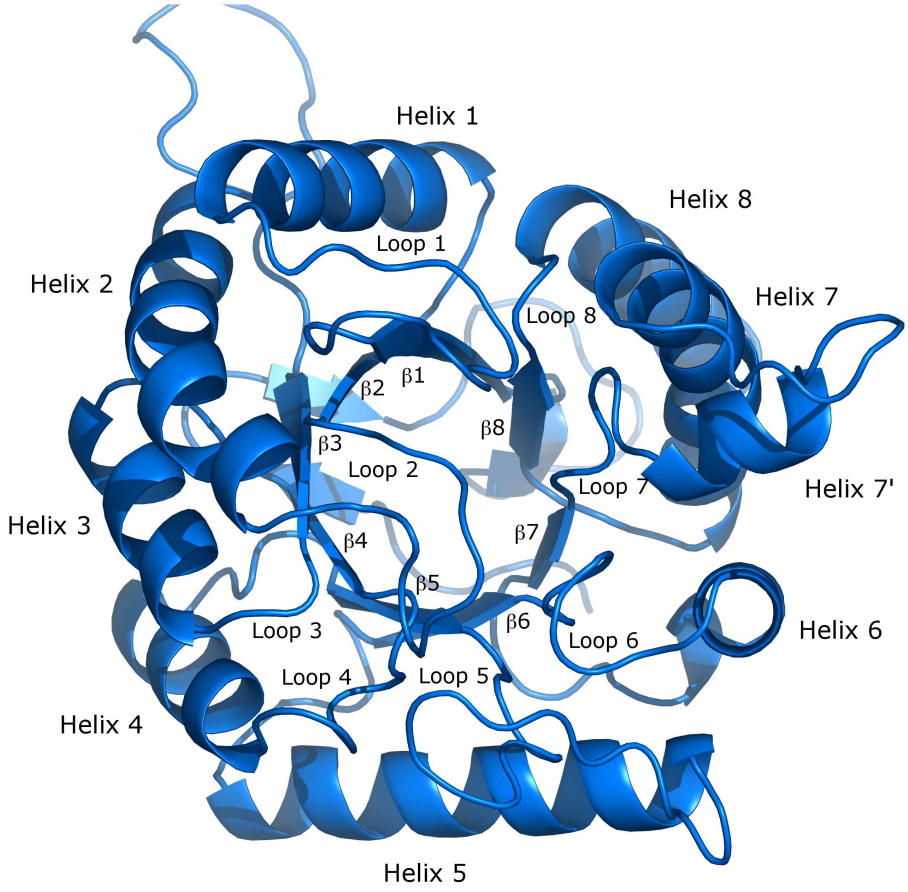

Figure 3.9: The modelled structure of *P. falciparum* DHPS. Insert 2 is situated just after helix 7'.

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.