Chapter 3

The Development of a Dynamic Receptor-Based Pharmacophore Model of *Plasmodium falciparum* Spermidine Synthase

3.1. Introduction to Pharmacophores

The concept of a pharmacophore was first coined by Paul Ehrlich in 1909 as "a molecular framework that carries (*phoros*) the essential features responsible for a drug's (*pharmacon*) biological activity" (Guner, 1999). The most widely accepted definition of a pharmacophore model was formulated by Peter Gund in 1977, which states that a pharmacophore model is "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity" (Gund, 1977). Pharmacophore models constitute pharmacophore features (PhF) which are moieties with specific chemical properties and a characteristic 3D geometry giving its biological activity (Milne *et al.*, 1998). The geometric arrangement of the PhFs complement the region of binding of the target and exerts the ligand's biological activity by interacting strongly with it. Pharmacophore models can be derived via analogue-based (ligand, indirect) or receptor-based (protein, direct) approaches (Guner, 1999; Dror *et al.*, 2004). The analogue-based approach uses ligands that have been experimentally shown to have activity against a target of unknown structure. In this approach the ligands are used to derive PhFs and construct pharmacophore models. The receptor-based approach uses a resolved ligand-target complex to derive PhFs and construct pharmacophore models.

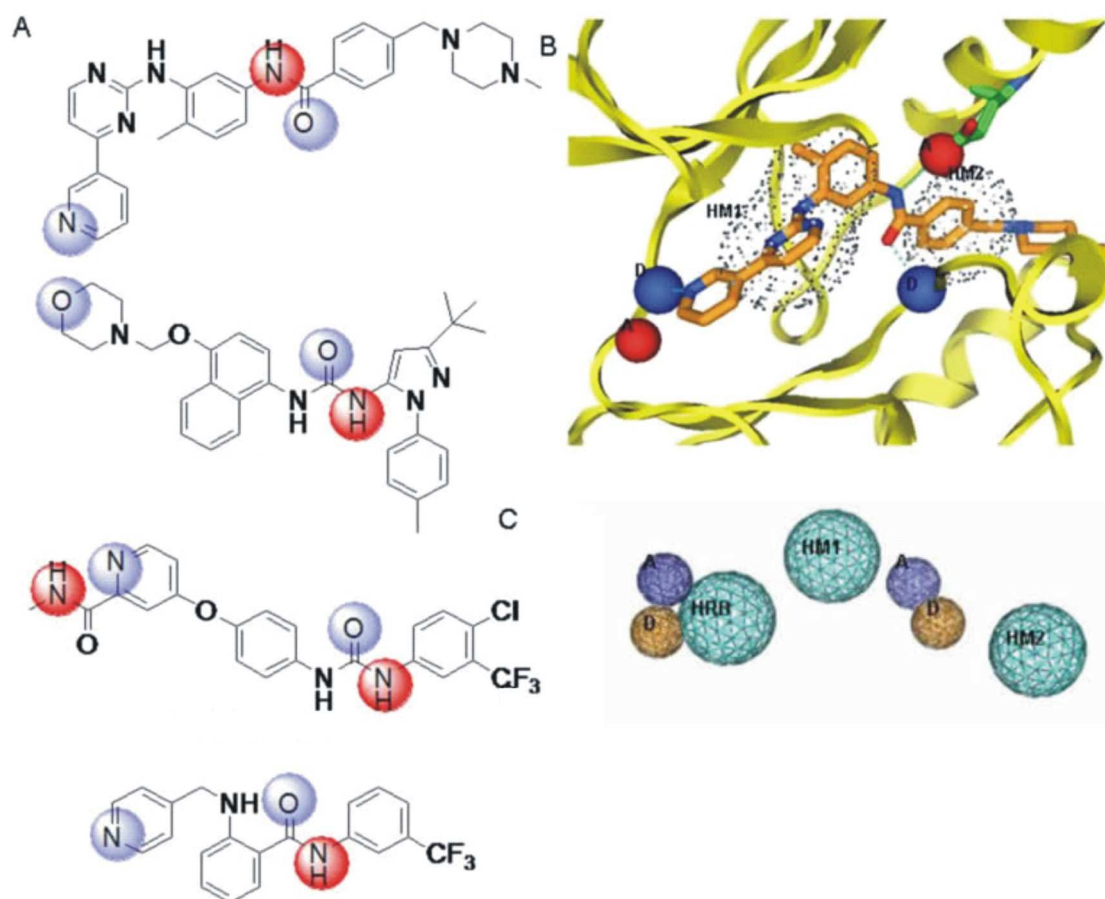These methods can be better understood by considering Figure 3.1.

Figure 3.1: An illustration to highlight the analogue-based and the receptor-based approaches from which a pharmacophore model can be derived. (A) Four ligands with biological activity against a target of interest which can be used in the analogue-based approach. (B) A resolved ligand-target complex illustrating the complementarity of the receptor to the ligand using PhFs. (C) Illustration of a pharmacophore model which could be derived from either the analogue-based or the receptor-based approach. (Adapted from the Gray Laboratories `http://research.dfci.harvard.edu/gray_lab/research.htm`)

Figure 3.1A describes an analogue-based approach represented by four ligands showing biological activity against a particular receptor. Common PhFs are highlighted (red and blue spheres) and constitute a pharmacophore model describing the most important interactions thought to be responsible for the ligand's biological activity. When deriving pharmacophore models from the structure-based approach a negative image of the active

site is constructed complementing the interactions between the receptor and ligand. This is illustrated in Figure 3.1B where the complementarity of the ligand and PhFs of the active site is shown. Figure 3.1C show the PhFs constituting a pharmacophore model, which could either have been derived from the known ligands (analogue-based approach) or from the ligand-target complex (receptor-based approach). Pharmacophore models therefore attempt to capture and describe both the chemical and geometric characteristics of a ligand that are the most important for binding to a target receptor. The receptor-based approach has the advantage that the correct geometry of the PhFs is known and can be used in more accurate screening of databases for new compounds. It is also possible to derive new PhFs from the target structure which is not possible using the analogue-based approach where only known inhibitors are used. The constructed pharmacophore models can then be screened against chemical databases to identify novel binders to a target of interest. The strength of the pharmacophore model approach therefore lies in its ability to predict/identify structurally diverse compounds representing a set of known chemical features (PhF) with a particular 3D geometry responsible for the biological activity against a target (Dror *et al.*, 2004; Khedkar *et al.*, 2007; Sun, 2008).

Pharmacophore models can additionally be used to perform scaffold-hopping and identify new ligands with different chemotypes but which still have a similar biological activity (Sun, 2008). Such a study was performed by Parkes *et al.* (2003) where they identified a novel structural class of inhibitors acting on influenza endonucleases. Palomer *et al.* (2002) used pharmacophore models to identify novel cyclooxygenase-2 selective inhibitors which differ with regards to their scaffold structures from the known inhibitors. Pharmacophore models have also been used in predictive strategies of ADMET properties and the *de novo* design of novel compounds (Ekins *et al.*, 2000; Dror *et al.*, 2004; Swaan and Ekins, 2005). The pharmacophore methodology also includes the use of pharmacophore finger printing and pharmacophore docking (Guner, 2002). Pharmacophore models are not only used in the drug discovery industry but also play an increasingly important role in the fragrance industry and are referred to as olfactophores (Guner, 2002). Having highlighted some of the most important uses of pharmacophore models it is not surprising

that the pharmacophore methodology has attracted much attention over the last few of years (Carlson *et al.*, 2000*a*; Dror *et al.*, 2004; Khedkar *et al.*, 2007; Guner, 2002; Sun, 2008). The importance of pharmacophore models in the current drug industry was further emphasized in a recent review by Khedkar *et al.* (2007) where it was reported that pharmacophore models are now being protected under intellectual property rights.

Comprehensive reviews have recently been published assessing the different pharmacophore approaches available (Dror *et al.*, 2004; Guner, 2002; Khedkar *et al.*, 2007; Sun, 2008). A general work flow diagram describing the steps involved in the development of a pharmacophore model using either the analogue or receptor-based approach is shown in Figure 3.2. The following section will detail the methodology and challenges associated with both the analogue and the receptor-based approaches used to derive pharmacophore models. These approaches will be discussed by referring to the steps listed in Figure 3.2.
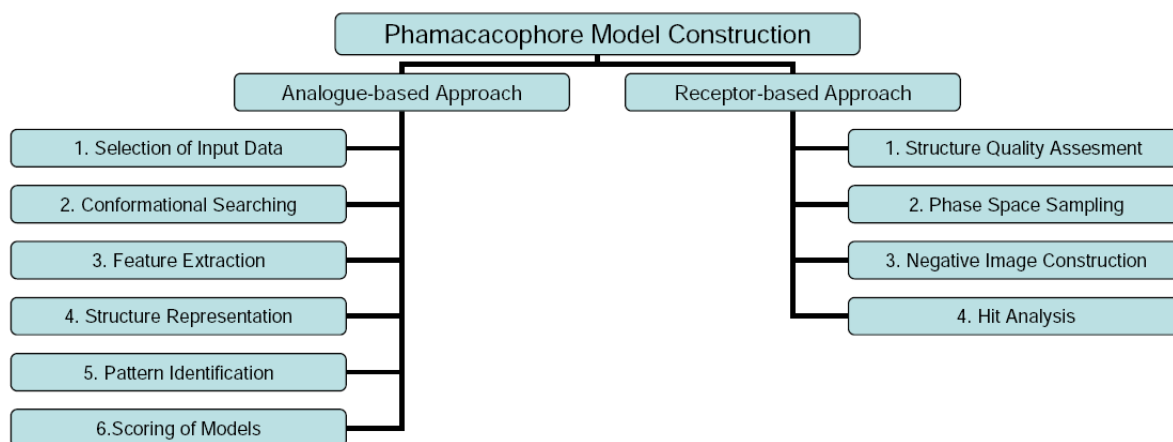
Figure 3.2: A diagram showing the general work flow procedure followed during the development of a pharmacophore model for both the analogue and receptor-based approaches. The receptor-based approach is described explaining the work flow for the development of a dynamic receptor-based pharmacophore model.

3.1.1. Analogue-based pharmacophore model

A six step frame work proposed by Dror *et al.* (2004) is shown Figure 3.2. As with most computational approaches the quality of the input data determines the quality of the output data. During the selection of the input data set, the type of ligands available,

the size of the data set and the chemical diversity of the compounds to be used need to be considered. The specificity and selectivity of the ligands for a specific target should also be considered to prevent the construction of inaccurate pharmacophore models. Therefore, when deriving analogue-based pharmacophore models, ligands with biological activity due to off-target effects should be avoided. The size and the chemical diversity of the ligands determine the software solutions which should be used and specify the confidence in the final derived pharmacophore models (Khedkar *et al.*, 2007). To generate a good pharmacophore model, the input data should have sufficient chemical diversity and consist of between 18-30 compounds (Kurogi and Guner, 2001). Larger sets of input data can also be used (1000-2000 compounds), but it is questionable how specific these compounds are since low quality pharmacophore models are obtained (Dror *et al.*, 2004).

The incorporation of conformational flexibility of ligands into pharmacophore models is crucial to and complicates the analogue-based pharmacophore model approach (Guner, 2002). The use of conformational searching (Figure 3.2) is essential to cover the conformational space of the ligands adequately, however this remains a major challenge (Dror *et al.*, 2004). Many good algorithms exist and include the polling function used in the Catalyst software suite (`http://www.accelrys.com`). During the conformational sampling the polling function penalize a newly generated conformation when it is too close to any other in the set, thereby ensuring the sampling of representative conformations (Smellie *et al.*, 1995).

Feature extraction (Figure 3.2) can be performed on three structural similarity levels and includes atom-based, topological-based and functional-based features (Dror *et al.*, 2004). The atom-based method defines a feature as a 3D position of an atom of a ligand while the topological-based approach groups atoms together and represents the feature as such i.e. a carbonyl group. The functional-based features represent the interactions thought to be important within the ligand-target complex and are usually represented by six functional groups including hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), positive ionizable features (PI), negative ionizable features (NI), aromatic

features and hydrophobic (HYD) features (Clement and Mehl, 2000). In the structure representation stage the pharmacophore features extracted are mapped together to form a representation of the ligand. The most commonly used method for mapping is graph theory (Dror *et al.*, 2004).

In the pattern identification stage (Figure 3.2), the features extracted from the different input ligand molecules are matched and pharmacophore candidates proposed (Khedkar *et al.*, 2007). A pattern can be described as a set of features having locations relative to each other in 3D space and is also frequently referred to as a configuration (Khedkar *et al.*, 2007). These patterns are used to search for ligands and it is thought that a ligand fits a pattern if it presents a set of features in a specific conformation and maps the ligand features to corresponding pharmacophore features. There are various algorithms to detect the embedded patterns in the 3D space represented by the ligand conformations generated for the input ligands. These algorithms include clique-detection, pairwise case, multiple case, exhaustive search and genetic algorithms (Dror *et al.*, 2004). However, the most popular means of identifying patterns given a set of ligands is the maximal common substructure (MCS) approach (Khedkar *et al.*, 2007). The MCS approach aims to identify the largest set of pharmacophore features embedded within the 3D space represented by the ligand conformations generated for the input ligands (Dror *et al.*, 2004). The main limitation of this approach is that it is based on the assumption that there is a single common pharmacophore model responsible for the observed activity (Dror *et al.*, 2004; Khedkar *et al.*, 2007). Barnum *et al.* (1996) proposed the use of a relaxed-MCS which allows for the generation of pharmacophore models where pharmacophore features need not be present in all the ligands but still need to meet a specific threshold. This therefore, allows the pharmacophore to miss a feature as long as it complies to a set threshold. The relaxed-MCS method has been shown to generate better models than the normal MCS method (Dror *et al.*, 2004).

The last stage of the analogue-based approach is the scoring system which is used to assess the quality of the pharmacophore model by using a scoring function that determines

the chance correlation that a ligand has given a certain pharmacophore model (Sutter *et al.*, 2000). To make the scoring of pharmacophore models more reliable, the scoring functions consider the number of PhF as well as their frequency of occurrence i.e. two negative ionizable features (features occurring less frequently) will be given a higher score than four hydrophobic features (occurring more frequently; Barnum *et al.* 1996). The newly generated pharmacophore models can then be used in screening.

3.1.2. Receptor-based pharmacophore model

The receptor-based pharmacophore approach (direct approach) should be followed if a 3D target structure is present. The receptor-based approach will be discussed using a dynamic receptor-based pharmacophore model (DPM) highlighting the differences with the normal receptor-based approach. The concept of a receptor-based dynamic pharmacophore model was first proposed by Carlson *et al.* (2000*a*). A dynamic pharmacophore model is in essence a pharmacophore model that attempts to account for the inherent flexibility of an active site and aims to reduce the entropic penalties associated with binding a ligand (Carlson *et al.*, 2000*a*). The biggest difference between the conventional approach and the DPM approach is the use of multiple protein conformations to incorporate receptor flexibility. The development of a DPM can be divided into four stages and will be discussed accordingly: 1) Protein structure quality assessment, 2) Phase space sampling, 3) Negative image construction and 4) Hit analysis (Figure 3.2).

3.1.2.1. Protein Structure Quality Assessment and pKa Predictions

The protein structure quality assessment stage can be subdivided into quality assessment of the target structure, active site identification and determining protonation states of the relevant ionizable groups. The integrity of the target structure is extremely important and therefore a thorough assessment of the structure is essential since it determines the conclusions that can be drawn from this approach and the confidence therein. A Protein 3D structure, may be described as a model of a protein structure that is derived and fitted to best represent experimental data (Laskowski, 2003). This implies that 3D structures will have errors associated with them, which can either be systematic or random (Laskowski, 2003). It is therefore of utmost importance that a thorough quality

assessment of the target structures be performed before embarking on the design of a DPM that is dependent on the quality of the input structure. The quality of a 3D structure can be assessed through analysis of various parameters including resolution, R-factor, R_{free}-factor, atomic B-factors and stereo-chemical parameters such as the Ramachandran plot (Brunger, 1992; Laskowski, 2003). The resolution of a 3D structure is the clearest measure of the apparent quality of a structure and serves as an indicator of the amount of detail that can be discerned from the computed electron density map and therefore the protein structure (Laskowski, 2003). It is generally known that a structure at 3Å resolution allows the tracing of a protein's backbone through the density map and that a structure with a 2Å resolution will allow accurate fitting of the side chains. Therefore, a crystal structure with a resolution of < 2Å may be used with a fair amount of confidence. The R-factor and R_{free}-factor are both measures that assess the fitting of atomic models to the observed diffraction data (Brunger, 1992).

The R_{free}-factor is deemed to be more reliable and less susceptible to manipulation during refinement (Brunger, 1992). Generally, structure models with R-factors < 0.2 and R_{free} -factors < 0.4 are considered to have good agreement with the experimental data and to be reliable. Atomic B-factors can be used as a measure of the precision of the coordinates of given atoms in a 3D structure. Generally, B-factors with values in excess of 40 indicate imprecise coordinates (Laskowski, 2003).

Protein 3D structures can additionally be assessed based on their stereo-chemical properties. The most common method used is the Ramachandran plot, which checks the stereo chemical integrity of the protein. The Ramachandran plot is essentially a scatter plot of the ψ (psi) versus the ϕ (phi) main-chain torsion angles of every amino acid of a protein which generally fall into favoured and unfavoured regions (Ramachandran, 1963; Laskowski, 2003). It is, however important to note that due to the nature of the data (statistical) from which the Ramachandran plot parameters have been derived, residues which fall outside the favoured areas are not necessarily wrong but may have crucial roles

in the function and structure of the protein.

The proper understanding of the physical properties of a protein are of importance in every aspect of modern biology. Although considerable biological data can be gathered and interpreted from the X-ray structure of a protein, little information regarding its physical characteristics is available (Nielsen and McCammon, 2003). Extensive calculations are thus needed to calculate these physical characteristics which may include ligand specificity, protein stability and the electrostatics of a protein (Nielsen and McCammon, 2003). The physical property of interest in this section is the protonation equilibrium of ionizable groups in proteins, which is described by a pKa value. Therefore detailed simulations of proteins, such as MD simulations, customarily require that the charge state of every ionizable group be specified *a priori* (Antosiewicz *et al.*, 1996). Knowing the ionization states of the active sites are helpful in the prediction of affinities of proteins for ligands (Antosiewicz *et al.*, 1996).

Protonation equilibria can be predicted by using the University of Houston Brownian Dynamics (UHBD; `http://adrik.bchs.uh.edu/uhbd/`) and YASARA (Krieger *et al.*, 2006) packages. The UHBD program uses the classical way of predicting the protonation equilibria by solving the Poisson-Boltzmann equation and by so doing, accounting for the dielectric constant of solute and solvent, as well as the ionic strength (`http://adrik.bchs.uh.edu/uhbd/`). YASARA uses an empirical equation to expresses the pKa calculations as a function of electrostatic potential, hydrogen bonds and accessible surface area by applying the particle mesh Ewald (PME) summation to simplify the interaction between ionizable groups and its environment (Krieger *et al.*, 2006). UHDB has two advantages over YASARA, being the treatment of implicit counter ions and applying different dielectric constants to the solvent and solute (Krieger *et al.*, 2006).

3.1.2.2. Phase Space Sampling

Phase space can be defined as all the possible states that a system can have in space. The aim of the phase space is to capture the dynamic behavior of a receptor and incorporate it at a later stage in the receptor-based pharmacophore model. Within the

structure-based drug design process the need to account for the dynamic behavior of a receptor has long been recognized as a complicating factor (Carlson and McCammon, 2000a). Historically, single high quality rigid protein structures (i.e. X-ray structures) have predominantly been used in the discovery process, the main reason being that the use of single rigid protein structures is a much faster process. However, the exponential increase of available computing power now provides the opportunity to incorporate protein flexibility into drug design. There are four major ways of incorporating protein flexibility into drug design; soft-docking, conformational sampling of the side chain in the receptor, generating a subensemble of structures and predicting loop flexibility and domain motions (Carlson and McCammon, 2000a). From the "lock-and-key" theory of ligand binding to a protein it is suggested that a protein exists in a single well-defined state with only one optimal complimentary state, however for this to be true the system would have to be very rigid. The energy landscapes of most proteins on the other hand have frequently been described by means of a folding funnel (conformational well) in which there are many highly unfavourable states that collapse via multiple routes into several possible favourable folded states (Ma *et al.*, 1999; Freire, 1999). The folded state of a protein is made up of a collection of structurally similar and nearly energetically equivalent conformations (Carlson and McCammon 2000a; Figure 3.3). This implies that a single structure might not be enough to describe the substates adequately, even when the weighted average of a crystal structure is used (Carlson and McCammon, 2000a). It is thus clear that protein flexibility has major implications for structure-base drug design and for the discovery of new lead compounds multiple protein conformations need to be considered.

A subensemble of protein structures can be either experimentally determined or computationally derived. Experimentally determined sources include X-ray crystallography and NMR structures. Computationally derived subensembles of structures can be obtained using Monte Carlo (MC) sampling (Verkhivker *et al.*, 2001) but have mainly been derived from MD simulations (Carlson *et al.*, 2000a; Deng *et al.*, 2005, 2006; Damm and Carlson, 2007; Damm *et al.*, 2008; Bowman *et al.*, 2007a). Considering experimentally

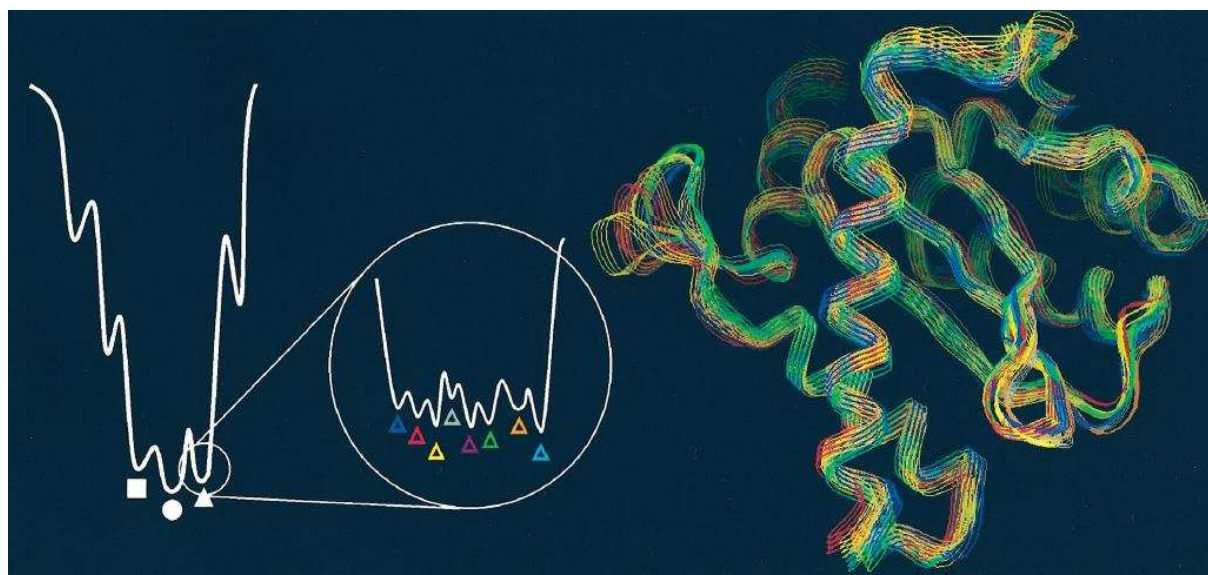Figure 3.3: A conformational funnel demonstrating multiple states for a protein. The flexibility inherent in a folded state (□ O△) is described by an ensemble of conformations (shown here as a collection of colored triangles (△)). A collection of structures from a MD simulation of HIV-1 integrase showing a sub-ensemble of states with a modest sampling of the backbone and a wide sampling of a small flexible loop on the right (Carlson and McCammon, 2000*a*). Expanding the minimum of a single state in the sub-ensemble (any of the colored triangles (△)) would reveal an additional series of sub-minima that arise from a variation in the orientation of the side chains. Adapted from (Carlson and McCammon, 2000*a*)

derived data, NMR structures, if available, are the preferred source of protein structures to represent the phase space, since it has been shown to sample a large portion of the conformational well. This technique, however is limited to proteins consisting of less than ~100 residues and is seldom applicable (Carlson and McCammon, 2000*a*). Multiple X-ray structures are another source of experimentally determined data and have been shown to better represent the phase space of a protein than MD simulation, but are system-dependent and is also not always available (Clarage *et al.*, 1995).

Considering computational approaches, Lamb and Jorgensen (1997) showed that free energy calculations such as MC sampling and MD simulations are the most rigorous computational methods for generating a subensemble of states and the calculation of their thermodynamic properties. These methods are reliable and are comparable with experimental data to within 1 kcal/mol (Lamb and Jorgensen, 1997; Carlson and McCammon, 2000*a*). Work done by Verkhivker *et al.* (2001) showed that when the MC

method is used for simulated annealing at temperatures ranging between 300K to 5000K, the system could still get trapped in local minima. This may be due to the difficulty of performing successful MC sampling on large flexible systems. The problem may, however, be overcome if some of the degrees of freedom are frozen or special methods are applied, but the drawback of these methods is a low phase space coverage (Leach, 2001). A comparison was drawn between the phase space sampled by 15 NMR structures of *E. coli* ribonuclease HI and the phase space sampled by a 1.7ns MD simulation, starting from a high resolution structure (PDBid 2RN2; 1.48Å; Philippopoulos and Lim (1999)). It was found that the MD simulation sampled similar structures to that of the NMR structures, however the sampling of the conformational well was not as wide. The NMR structures also showed more flexibility in the side chains and backbone.

It is therefore clear that MD simulations are the preferred computational method to sample the conformational well of the phase space for the target structure of interest and this approach was selected accordingly for this study. Molecular dynamics can be defined as the integration of Newton's equations of motion on a set of atoms allowing for the generation of a successive series of configurations resulting in a trajectory that specifies how the positions and velocities vary over time (Leach, 2001). The numerical integration of Newton's laws of motion reveals the intrinsic motions of the system under the influence of an associated force field (Schlick, 2002). This force field is representative of a functional form, which is used to describe the energy of a system as a function of the nuclear position of the atoms and is governed by a set of predefined parameters (Leach, 2001). MD simulations thus combine both the spatial and temporal aspects of conformational sampling (Schlick, 2002).

The sampling of the phase space of a protein results in a large number of similar structures, which are captured during a MD simulation. These data needs to be processed and analyzed to obtain a subensemble of structures representative of the covered phase space. This can be done by using an unsupervised method known as clustering. Clustering deals with the intrinsic grouping of unlabeled data sets. In other words clustering deals

with large amounts of data, which are grouped or clustered together based on similar properties from which a smaller representative data set can be selected (Leach, 2001; Causton *et al.*, 2003). Clustering is therefore a good method for finding a smaller set of representative structures of the phase space sampled during the MD simulation.

3.1.2.3. Negative Image Construction

The negative image construction stage involves the identification of chemical features within the active site and the derivation of the relationship of their spatial geometry through the construction of DPMs. The identified chemical features are represented by PhFs. The PhFs identified from the negative image of the active site are thought to provide complementary PhFs which can be used in the identification of strong binding ligands.

This stage can be broken down into molecular interaction field (MIF) analysis and the identification of exclusion volumes (EV; Figure 3.4). From MIF analysis PhFs can be identified which in turn can be combined with the identified EVs and used to construct DPMs (Figure 3.4). This section will be discussed using the flow diagram in Figure 3.4.

The conventional way of constructing a negative image is by generating an interaction map of the active site by identifying chemical features such as HBDs, HBAs and HYD features followed by the placement of complementary features within reasonable chemical space (Dror *et al.*, 2004; Bohm, 1992, 1994). Negative image analysis has also been performed by flooding the active site with chemical probes followed by their optimization and feature extraction, these features are often described as binding hotspots (Carlson *et al.*, 2000*a*; Damm and Carlson, 2006). The latter method can also be applied in the construction of DPMs by flooding the subensemble of structures identified during phase space sampling (Carlson *et al.*, 2000*a*; Damm and Carlson, 2006). The binding hotspots identified using these methods can be described as binding areas with specific chemical features (i.e. HBA, HBD and HYD), which are present in all subensemble structures.

A third method, MIF analysis, is the preferred method for this study and is used to
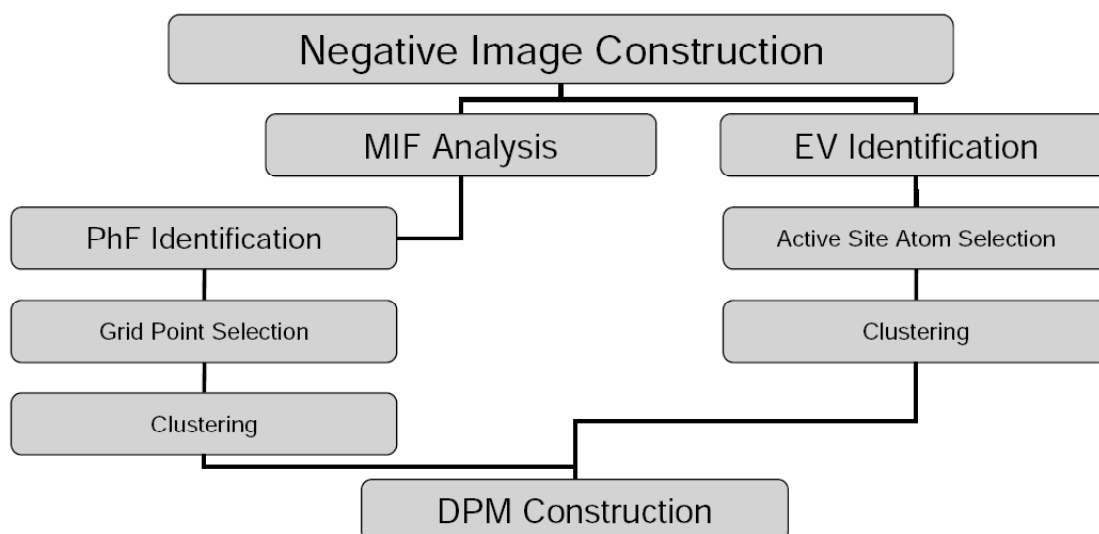
Figure 3.4: A work flow of the negative image construction stage of the development of a DPM. This stage can be broken down into MIF analysis and EV identification which are combined after completion to construct DPMs. MIF analysis can further be broken down into PhF identification, which involves grid-point selection and subsequently their clustering to obtain the representative positions of the PhFs. The EV identification involves the identification of active site atoms to be used as EVs and subsequently their clustering to get the representative positions. The identified PhFs and EVs can then be combined to generate DPMs.

obtain the coordinates in Cartesian space of energetically favourable binding sites within the active site of the target protein and are represented by PhFs. This is done using a program called GRID (`www.moldiscovery.com`). The grid-based method uses specific non-covalent interactions between a molecule of known three dimensional structure (the "Target"), and a small chemical group (the "Probe") to predict interactions. The approach can be used to study a variety of targets and includes enzymes, nucleic acids, poly-saccharides, glycoproteins, peptides, membranes, crystals, drug molecules, photographic materials, dyes and many other organic chemicals (`www.moldiscovery.com`).

The basic principle of grid-based methods entails that an electrostatic potential is calculated around the protein or target. This electrostatic potential is calculated based on the x, y and z coordinates of the atoms of the target after which the model is surrounded by an imaginary orthogonal grid (Figure 3.5). Subsequently, starting at the first grid point the electrical potential is calculated at that particular point for a unit

electrostatic charge (which is brought from infinity; for more information see Cruciani (2006)). This is done for all the grid points and is tabulated giving their coordinates and electrical potential. Electrostatic potentials do not normally allow the differentiation between favourable binding sites for a primary, a secondary, a tertiary amine cation, a tetramethyl ammonium, a peridium or a sodium cation. However, the grid-based method is an attempt to compute analogue potentials which have chemical specificity within the active site (Cruciani, 2006). This accumulation of chemical specific information contributes to the understanding of interaction between the target and its molecules.
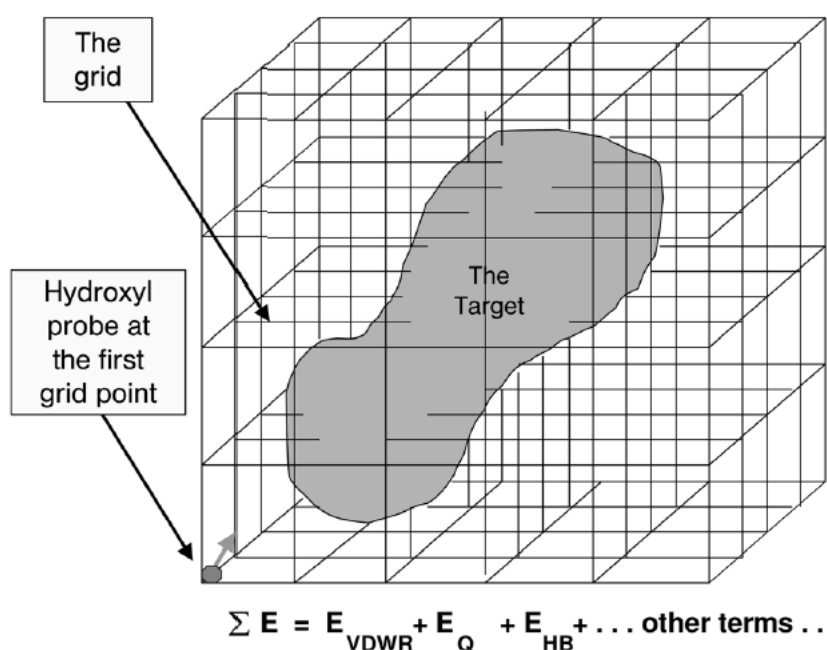
Figure 3.5: The setup of MIF analysis using a grid-based method. An imaginary orthogonal grid constructed around the target is shown. The electrical potential for the probe is then calculated starting at the first grid point as can be seen in this figure. Adapted from Goodford (2005).

The grid-based method has three features that distinguishes it from normal programs calculating the electrostatic potentials (Goodford, 2005): 1) GRID probes are often anisometric, 2) the target responds when the probe is moved around it from place to place and 3) it is assumed that both the target and the probe are immersed in water. Most of the GRID probes are anisometric since probes can represent either an atom or a small group of atoms (Goodford, 2005), for example a carbonyl oxygen probe, which consists of

an oxygen atom with two lone pairs (sp^2) represents a probe that has size, a polarizability and an electrostatic charge with both the lone pairs being able to form hydrogen bonds. During calculations this oxygen atom is placed at the center of a grid-point, bad contacts are identified and nearby HBDs are searched for. The probe is then rotated around the grid-point to find the best possible interactions at that specific grid-point. The procedure is performed for all the grid-points of the grid. The use of anisometricity provides the user with chemical specific information of the target, which is more informative than just an electrostatic potential map (Cruciani, 2006). When a probe (e.g. hydroxy probe sp^2) is positioned at a specific grid-point the orientations of the hydrogen will be in a random position. GRID allows the target to be responsive which would mean that residues within the vicinity of the probe can be adjusted to form the most favoured interactions (Figure 3.6). Another advantage of the grid-based method is that it is able to treat the target as if it is present within a biological system, submerged in a water solution containing different chemical entities influencing the dielectric constant. Most methods uses a dielectric constant of a vacuum (considered to be one), however GRID deals with this problem by treating the outside of proteins with a dielectric constant of 80 and the inside with a dielectric constant of 4.

MIF analysis using the grid-based method is performed with various different probes to explore the active site and its affinity for binding different chemical groups. Depending on the active site three different classes of probes are usually used to identify the PhFs. These include HBA, HBD and hydrophobic (includes aromatic groups) probes. Depending on the system and the user there are various different probes which can be used to identify these PhFs. MIF analysis results in grid-points having grid values (energy values) which are extracted, clustered together and used to produce the PhFs.

The construction of a negative image for the active site of an receptor results in a large number of PhFs. A single query or DPM containing all PhFs would be too complicated to retrieve hits from a database search (Dror *et al.*, 2004; Venkatachalan *et al.*, 2000). Therefore it is important to construct multiple DPMs which can be used
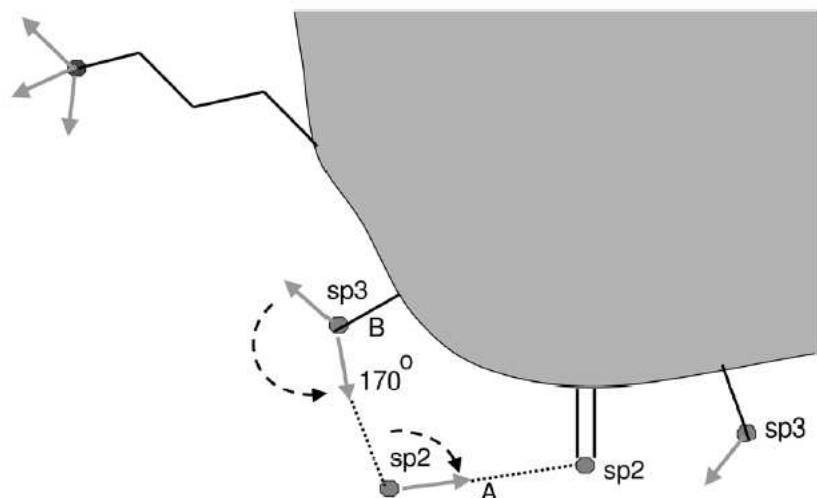
Figure 3.6: An illustration of the ability of target to respond to the introduction of a chemical probe in its active site. Here a hydroxy probe (sp^2) is introduced at a grid point having a random orientation to the target, it forms a hydrogen bond with a hydroxyl group of a Ser residue labeled B. The hydroxyl probe can be adjusted to make a hydrogen bond between its hydrogen atom and the the sp^2carbonyl group (labeled A). The hydroxyl group of Ser (labeled as B) can then respond to this change and be oriented in such a way that it forms a hydrogen bond with the probe. This then results in a more favourable interaction. The grey shaded area is representative of the target. Adapted from Cruciani (2006).

to search the chemical databases and retrieve compounds fitting these DPMs. A set of guidelines for the selection of PhF to construct pharmacophore models have been proposed by Venkatachalan *et al.* (2000). They proposed that a query or pharmacophore model should consist of between three and seven PhFs. It is also suggested that if there are N HBA identified, assuming one constructs queries (DPM) with n acceptors (n<N), then there are N!/[n!(N-n)!] combinations to consider. This needs to be done for each feature type (Venkatachalan *et al.*, 2000). This however, might be too large a number of DPM to be searched and it is suggested that knowledge from known inhibitors and mutation studies be considered when selecting PhFs for the construction of DPMs. The inter-feature distance should be considered in the selection of PhF to construct DPMs since large inter-feature distances result in the retrieval of large molecules which is not always desirable. Venkatachalan *et al.* (2000) suggested a minimum inter-feature distance of 1.5Å to avoid overlap between the features.

Even when following these guidelines a large number of hits can be returned from database screening, therefore exclusion volumes are used to make the searches more specific for the target molecule. The use of a receptor-based pharmacophore model has the advantage of exploiting structural information such as shape and volume which can be derived directly from the active site (Dror *et al.*, 2004). The shape and volume are represented by exclusion volumes and should be added to the pharmacophore models. These exclusion volumes serve as a barrier preventing the overlap of ligands with the receptor atoms making the database screens more specific (Venkatachalan *et al.*, 2000). Exclusion volumes are usually derived by selecting atoms within a certain distance of the active site. The exclusion volumes delimit the space accessible to the ligands within the active sites since the exclusion spheres act as hard spheres which do not allow for mapping during database searches (Hoffman *et al.*, 2000).

3.1.2.4. Hit Analysis and *In Vitro* Testing

This stage involves the screening of DPMs generated in the previous stage against a chemical database of choice. The hits identified are filtered, evaluated and the best scoring compounds selected for *in vitro* testing. A detailed breakdown of this stage is provided in Figure 3.7.

The hit analysis stage starts by the selection of either a commercially or publicly available chemical databases to be screened (Oprea and Tropsha, 2006). One of the biggest challenges of virtual screening today is the cost and ready availability of databases for the use in screening, since they are usually difficult to prepare and not curated (Irwin and Shoichet, 2005). Large efforts have been made over the last few of years to generate chemical databases meeting these challenges. These efforts include the Zinc - is not commercial (ZINC) database (Irwin and Shoichet, 2005), National Cancer Institute (NCI) (`http://cactus.nci.nih.gov/`), PubChem (`http://pubchem.ncbi.nlm.nih.gov`), the Super Drug DataBase (Goede *et al.*, 2005), the Drug Bank (Wishart *et al.*, 2008) and the SuperNatural database (Dunkel *et al.*, 2006).

These databases are not all freely available for download and screening but are avail-
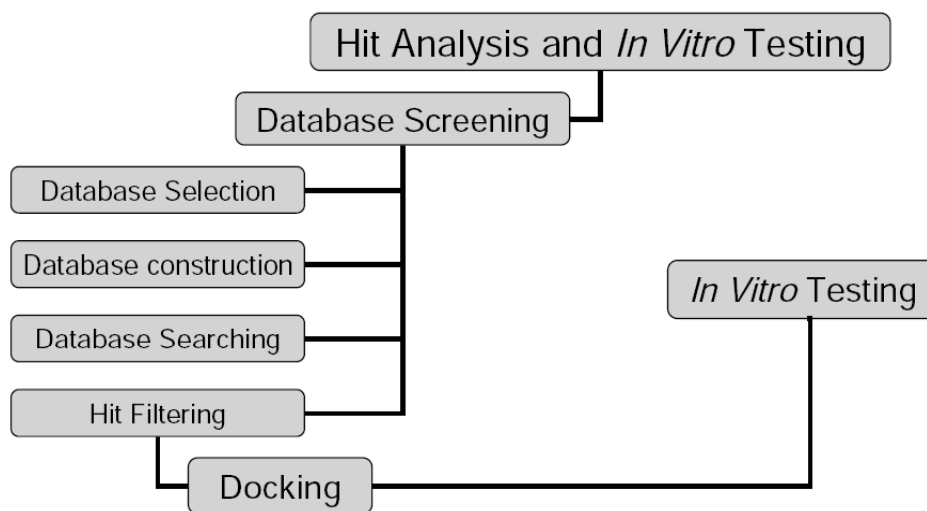
Figure 3.7: A work flow of the hit analysis and *in vitro* testing stage of the development of a DPM. This stage starts with database screening which involves database selection, database construction, database searching and hit filtering. The filtered compounds are subsequently docked and selected for *in vitro* testing.

able on-line for similarity searches. Irwin and Shoichet (2005) suggested that the "gold standard" for docking databases in academia are the commercially available, Available Chemical Database (ACD; `http://www.mdli.com`) having 250 000 purchasable compounds and over 2.3 million compounds in the ACD-SC (screening compound set). The Cambridge Structural Database (CSD, `http://www.ccdc.cam.ac.uk`, 430 000 compounds) and the ChemNavigator database (`http://www.chemnavigator.com/` over 10 million compounds) are some other commercially available databases often used in virtual screening. These databases have the drawback that they leave the user with the challenges of deciding on the protonation states, charges, tautomeric forms, and removal of salts (Irwin and Shoichet, 2005). These challenges are not unique to commercial products but are also found in most publicly available databases such as the Ligand-Info database (`http://ligand.info/`). The ZINC database containing over 8 million purchasable compounds is the first database where all the above-mentioned aspects have been addressed by the curators. The ZINC database has numerous sub-sets available to be downloaded. These include subsets such as lead-like, drug-like, fragment-like, Verneralis-filtered, all purchasable compounds, *etc.* (`http://zinc.docking.org/`). In the present study the

drug-like subset of the ZINC database, which was pre-filtered by applying Lipinski's rule of five (Lipinski, 2000). Lipinski's rule of five can be defined as follow: a compound should not have more than five hydrogen bond donors, no more than ten hydrogen bond acceptors, have a molecular weight less than 500 dalton and have a partition coefficient (log P) less than five.

The selection of the database is followed by its construction and subsequently the searching of the DPMs. These searches return compounds with a fit value which can be described as a measure of how well a compound fits a pharmacophore model. These fit values are used to rank and filter compounds. The best-fitting compounds are considered to have the most potential of showing biological activity. Docking is then used as a complementary tool and aims to prioritize compounds identified during pharmacophore screens for biological assays.
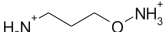
Molecular docking can be defined as a computational tool, which aims to identify the correct binding poses of ligands to the binding pocket of a specific target and to predict its affinity (Krovat *et al.*, 2005). Many docking programs are available and include software such as DOCK (Kuntz *et al.*, 1982), GOLD (Jones *et al.*, 1997), AutoDock (Goodsell *et al.*, 1996), FlexX (Rarey *et al.*, 1996), LigandFit (Bissantz *et al.*, 2000), Glide (`www.schrodinger.com`), CDocker (`http://www.accelrys.com`) and ICM (`http://www.deltahpc.com`). Many comparisons between the different docking packages have been made with the most popular docking programs being used including AutoDock, GOLD, FlexX DOCK and ICM (Sousa *et al.*, 2006). It is difficult to compare docking programs with each other since it is widely known that accuracy of docking programs varies significantly from target to target (Sousa *et al.*, 2006; Schulz-Gasch and Stahl, 2003; Bissantz *et al.*, 2000). Molecular docking is at a mature stage of development, but still has many challenges to overcome. The accuracy of most docking programs to dock ligands to targets varies between 1.5 to 2 Å with reported success rates in the range of 70-80% (Halperin *et al.*, 2002; Sousa *et al.*, 2006; Bursulaya *et al.*, 2003). Gohlke and Klebe (2002) commented that significant improvement on this range is unlikely at

present even with the inclusion of receptor flexibility. The major drawback of docking methods at present is the lack of reliable scoring functions. The aim of a scoring function is the identification of the correct binding poses of a ligand through its lowest energy, and the ranking of protein ligand-complexes according to their binding affinities (Krovat *et al.*, 2005). The scoring functions can be divided into three groups: 1) empirical scoring functions, 2) force field-based scoring functions and 3) knowledge-based scoring functions (Krovat *et al.*, 2005). Consensus scoring is generally used in docking to account for the differences between the scoring functions.

Docking results can be evaluated based on the docked poses and their corresponding docking energies. It has long been known that docking to ensembles of protein structures results in significant improvements over single structure docking (Knegtel *et al.*, 1997). As was discussed earlier, proteins do not exist in single rigid conformations and the binding of ligands often induce significant changes to the active site. Therefore docking to multiple target structures improves the selection of compounds which shows the best potential to have biological activity (Knegtel *et al.*, 1997).

This concludes the explanation regarding the development of a receptor-based DPM. This process is an iterative one and information gained from any of the four stages can be incorporated to optimize the DPM model (Figure 3.2).

Reference should be made to substructure searches to identify similar compounds, which can be used as a complementary tool to the development of a DPM. Known ligands can be screened against various on-line databases in particular the SDD, DrugBank and PubChem. Similarity searches of compounds identified during the pharmacophore screens can be used to search other databases, since no database contains all compounds available. SciFinder can also be used to perform similarity searches of commercially available compounds (Wagner, 2006). The similarity searches can be used to identify scaffold structures of the compounds identified during the pharmacophore screening. These scaf-

folds may show more desirable characteristics giving the modeler the ability to select compounds to evaluate specific characteristics on a substructure of an identified scaffold.
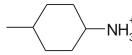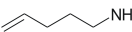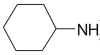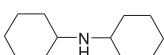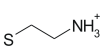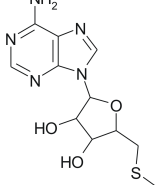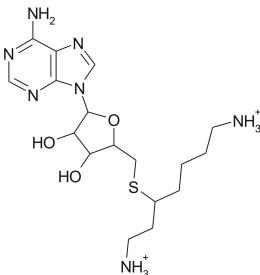
3.1.3. Knowledge of known inhibitors

Prior to the discussion of MIF analysis, pharmacophore models, database screening and docking, a summary is given providing information of known inhibitors and their interactions with PfSpdSyn. Table 3.1 provides the names, structures, enzyme IC_{50} and growth IC_{50} values available in the literature. These inhibitors include 4MCHA, 5-amino-1-pentene (APE), cyclohexylamine (CHA), APA, dicyclohexylamine, 2-mercapto-ethylamine, MTA and AdoDATO. The inhibitors tested for their effect on spermidine metabolism have been discussed in section 1.2.3.

It is important to note that cyclohexylamine identified by Hibasami *et al.* (1980) as a spermidine synthase inhibitor was incorrectly referred to as dicyclohexylamine but was corrected in a letter to Journal of Biochemistry by Batchelor *et al.* (1986). Inhibitors binding to the putrescine binding cavity include 4MCHA, APE, cyclohexylamine and APA (Table 3.1). The most potent inhibitor of PfSpdSyn is *trans*-4MCHA (Table 3.1). 4MCHA binds within the putrescine binding cavity and has been co-crystallized with PfSpdSyn (PDBid 2PT9; Dufe *et al.*, 2007). It is noteworthy to mention that the crystallization of 4MCHA within PfSpdSyn could only be achieved in the presence of dcAdoMet (Dufe *et al.*, 2007). This is suggestive of two things, first that 4MCHA competes for binding with putrescine and second, that cooperative binding exists between 4MCHA and dcAdoMet. Superimposing of the crystal structure from PfSpdSyn co-crystallized with 4MCHA and dcAdoMet (PDBid 2PT9) and that of the HsSpdSyn co-crystallized with putrescine and MTA (PDBid 2O06) revealed that the 4MCHA and putrescine binding modes are very similar (Figure 3.8). 4MCHA and putrescine lies in the same plane with the butyl moiety of putrescine aligned with the hexyl ring of 4MCHA. This is in agreement with results published by Dufe *et al.* (2007). The amine group of 4MCHA aligns with the non-attacking nitrogen of putrescine whereas the methyl group of 4MCHA aligns with the attacking nitrogen of putrescine (Figure 3.8). 4MCHA forms a hydrogen bond between its

Table 3.1: Inhibitors tested *in vitro* on PfSpdSyn. It should be noted that this table includes only inhibitors for which PfSpdSyn IC$_{50}$ values tested *in vitro* were available. When available the growth IC$_{50}$ values of the selected inhibitors are included. The discussion of inhibitors represented in this table is broken down to inhibitors binding in the putrescine binding cavity (PBC), the dcAdoMet binding cavity (DBC), the whole of the active site (WAS) and inhibitors without known binding cavities (UBC).

Inhibitor	Structure	Enzyme inhibition IC$_{50}$ μM	Growth inhibition IC$_{50}$ μM	Binding cavity
4MCHA		1.4 ± 0.1 [*]	34.2 ± 4.0 [*]	PBC
APE		6.5 ± 2.1[*]	83.3 ± 3.3[*]	PBC
Cyclohexylamine		19.7 ± 3.1[*]	198 ± 47 [*]	PBC
APA		84 ± 21[*]	1.0 ± 0.3 [*]	PBC
Dicyclohexylamine		>1000[*] (47.44)[♯]	342 ± 57 [*]	UBC
2-Mercaptoethylamine		76 ± 10[*]	254 ± 42 [*]	UBC
MTA		159 ± 27[*]	N.D.	DBC
AdoDATO		8.5 ± 0.3 [**]	N.D.	WAS

[*]Haider *et al.,* 2005
[**]Dufe *et al.,* 2007
[♯]Moritz *et al.,* 2004

amine group and Asp 199 (Figure 3.9), which corresponds to the hydrogen bond formed between Asp 176 and the non-attacking nitrogen of putrescine in the HsSpdSyn (PDBid 2O06).

A comparison of the negatively charged cavity where the attacking nitrogen of putrescine binds (catalytic center) was made between the HsSpdSyn (PDBid 2O06) and
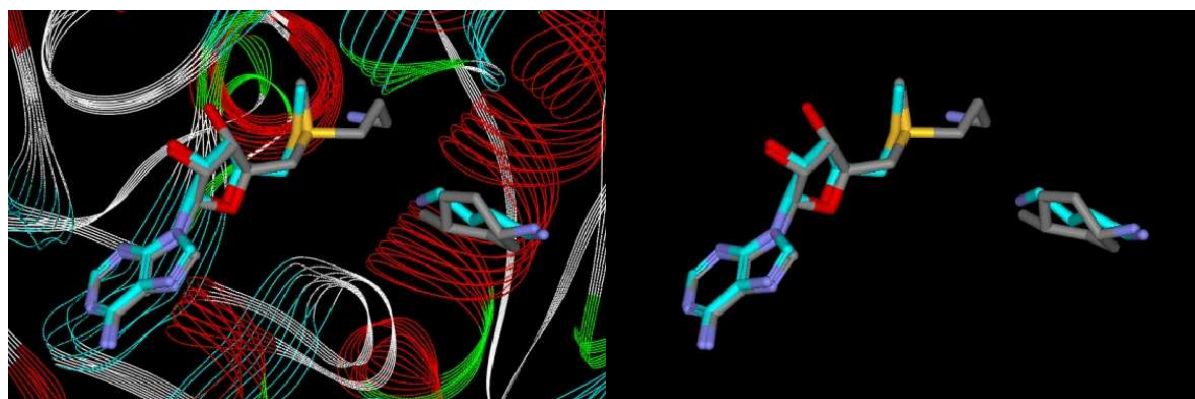
Figure 3.8: Superimposed structures of 4MCHA (grey colored carbons) and putrescine (cyan colored carbons) from PfSpdSyn (PDBid 2PT9) and HsSpdSyn (PDBid 2O06) . The structures of dcAdoMet (grey coloured carbons) from the PfSpdSyn and MTA (cyan colored carbons) from the HsSpdSyn structures are also displayed. The image on the left includes the protein backbone as ribbons whereas the picture on the right only displays the ligands.

PfSpdSyn (PDBid 2PT9) crystal structures to assess the changes, which occur upon binding of putrescine and 4MCHA. An overlay of the crystal structures from HsSpdSyn and PfSpdSyn containing their respective ligands was performed and it was found that most residues are remarkably similar in orientation (Figure 3.10). Dufe *et al.* (2007) reported that the methyl group of 4MCHA does not occupy the same 3D space as the attacking nitrogen of putrescine, however it can be seen in Figure 3.10 that it does. The methyl group of 4MCHA was within 0.1Å of the attacking nitrogen of putrescine. The most significant conformational change which exists between the HsSpdSyn and the PfSpdSyn was found to be Ser 198 and is illustrated in Figure 3.10.

Cyclohexylamine, one of the first inhibitors of PfSpdSyn has an enzyme IC_{50} value of 19.7μM and a growth IC_{50} of 198 μM. Cyclohexylamine differs from 4MCHA only by a methyl group at the para position. The inclusion of the methyl group (in the *trans* conformation) reduces the enzyme IC_{50} value to almost a 20th of that of cyclohexylamine.

APE is very similar in structure to 4MCHA and a molecular overlay is shown in Figure 3.11. APE has an enzyme IC_{50} value of 6.5μM and a growth IC_{50} of 83μM (Haider *et al.*, 2005) making it the second best known inhibitor of PfSpdSyn at present. No additional information on its mode of inhibition is available and it is assumed to interact in a similar
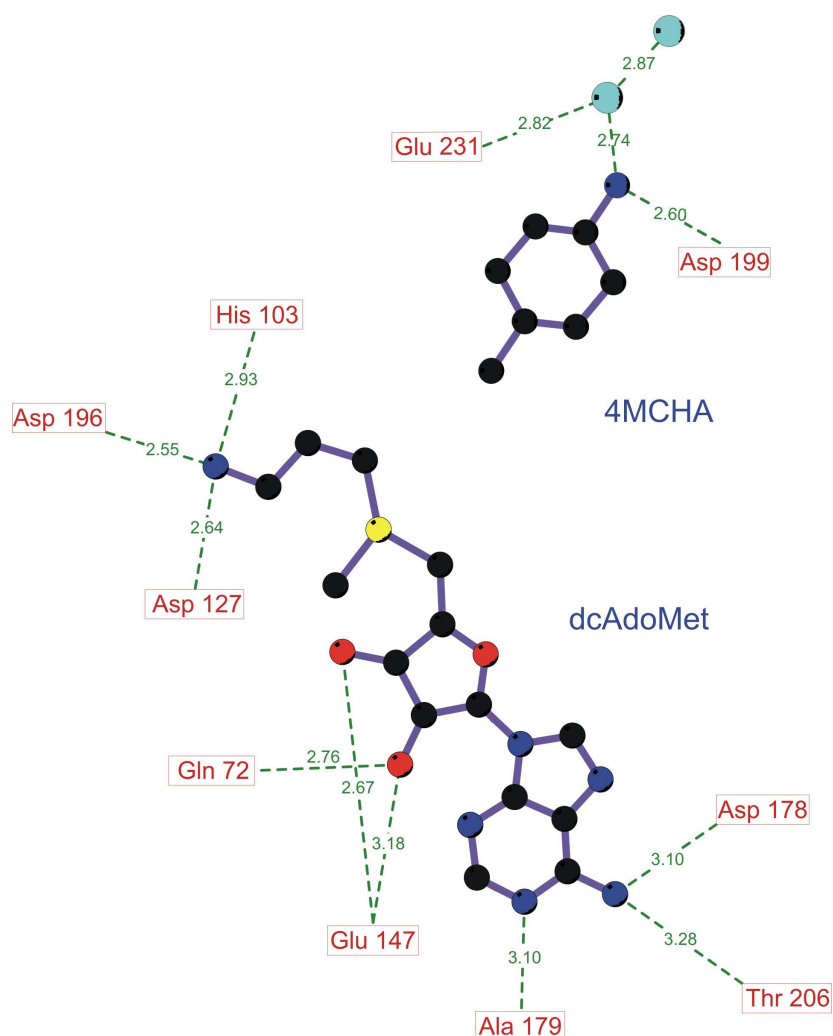
Figure 3.9: A diagram showing hydrogen bond formation between PfSpdSyn and the ligands dcAdoMet and 4MCHA co-crystallized (PDBid 2I7C). Generated using Ligplot (Wallace *et al.*, 1995).

manner to 4MCHA.

A molecular overlay of APA with 4MHCA can be seen in Figure 3.11, without implying that APA binds in this conformation or pose. APA, originally tested and designed for the inhibition of ODC, has an enzyme IC_{50} value of 84 μM for PfSpdSyn and an IC_{50} of between 3 and 8 nM for PfODC (Gupta *et al.*, 2005). APA has a growth IC_{50} of 1.0 μM when tested against *P. falciparum* cultures (Gupta *et al.*, 2005; Haider *et al.*, 2005), the effect of APA on *P. falciparum* cultures can thus not be explained as a result of inhibition of a single protein but probably the inhibition of multiple proteins. No
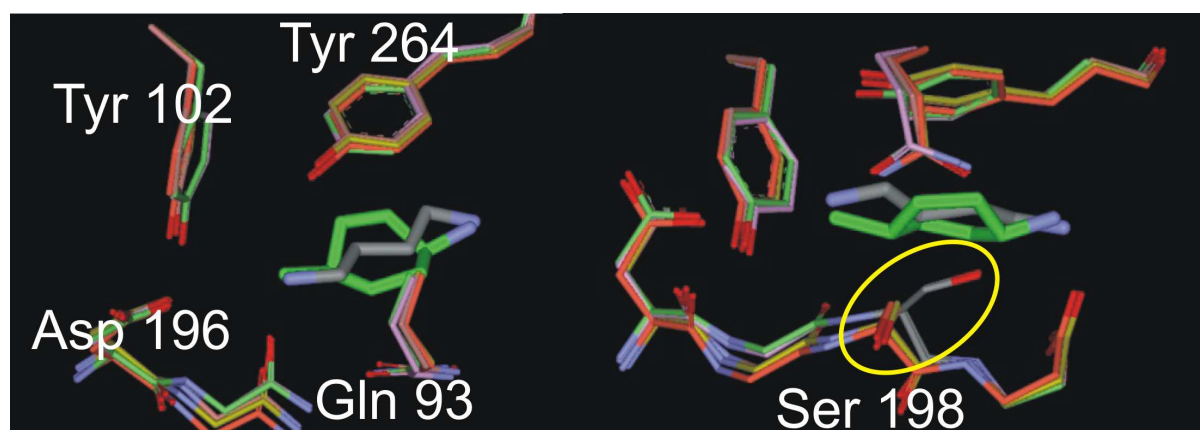
Figure 3.10: A comparison between the catalytic center of PfSpdSyn and HsSpdSyn after an overlay of the respective crystal structures. In the center of the illustrations are 4MCHA in green and putrescine in CK colors. *(Left)* An illustration of the similarity between the binding cavity of the attacking nitrogen of putrescine of the HsSpdSyn and PfSpdSyn crystal structures. *(Right)* An illustration of the biggest conformational change between the HsSpdSyn and PfSpdSyn crystal structures, Ser 198, encircled in yellow.
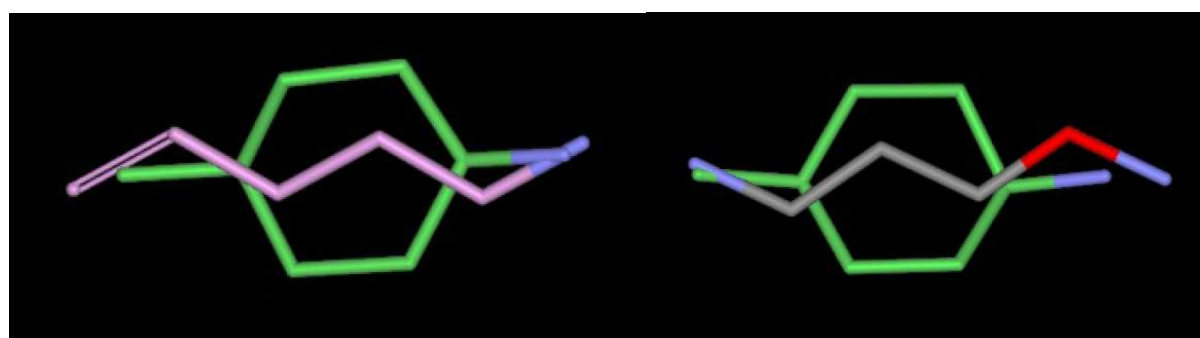

Figure 3.11: A molecular overlay of similar inhibitors of PfSpdSyn. *(Left)* A molecular overlay of APE and 4MCHA. *(Right)* A molecular overlay of APA and 4MCHA.

additional information during the current study was available on the mode of inhibition of APA in PfSpdSyn.

As discussed earlier dicyclohexylamine is not an inhibitor of PfSpdSyn. However, Moritz *et al.* (2004) reported an enzyme IC$_{50}$ of 47 μM in contrast with results from Haider *et al.* (2005) which showed an IC$_{50}$ value of > 1000 μM. It might be that Moritz *et al.* (2004) used the same product as Hibasami *et al.* (1980), however this could not be concluded from their article.

In a study by Shirahata *et al.* (1991) they explored both the putrescine and spermidine binding cavities of the respective aminopropyltransferases in the mammalian SpdSyn for possible inhibitors. A set of 25 cyclohexylamine-related compounds (Table 3.2) and a set of 13 alkylamines (Table 3.3) were tested and their respective IC_{50} values determined. In

Table 3.2: Inhibition of spermidine synthase activity by cyclohexylamine-related compounds of the mammalian SpdSyn (Adapted from Shirahata *et al.* 1991).

No	Structure	IC_{50} (μM)	% Inhibition at 1mM	No	Structure	IC_{50} (μM)	% Inhibition at 1mM
1		8.1		14		68	
2		103		15		>1000	33.2
3		>1000	40.4	16		>1000	8.2
4		>1000	37.0	17		>1000	<1.0
5		300		18		19	
6	(cis)	430		19		15	
7	(trans)	1.7		20		30	
8		330		21		>1000	18.5
9		>1000	29.1	22		>1000	<1.0
10		108		23		>1000	<1.0
11		>1000	<1.0	24		5.5	
12		>1000	<1.0	25		107	
13		>1000	<1.0				

addition 17 diamine compounds (Table 3.4) were tested for potential substrate activity (putrescine-like).

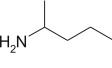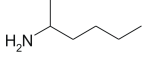Table 3.3: Inhibition of mammalian SpdSyn activity by alkylamines
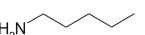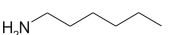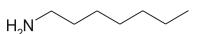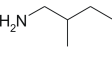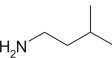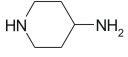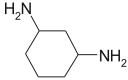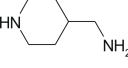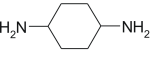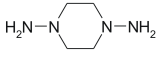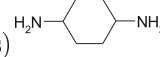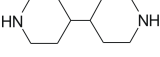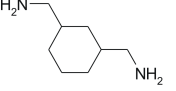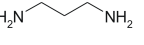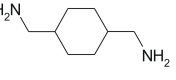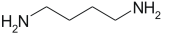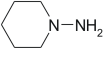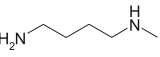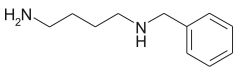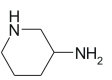
No	Structure	IC$_{50}$ μM	No	Structure	IC$_{50}$ μM
26		35	33		450
27		3.8	34		150
28		3.6	35		>1000
29		104	36		45
30		>1000	37		20
31		1000	38		7.8
32		250			

Table 3.4: Screening of potential diamine substrates of the mammalian SpdSyn.

No	Structure	%MTA production	No	Structure	%MTA production
39		<1.0	48		<1.0
40		<1.0	49		2.7
41	(cis)	3.0	50		<1.0
42	(trans)	11.6	51		<1.0
43		<1.0	52		<1.0
44		<1.0	53		100
45		<1.0	54		18
46		<1.0	55		<1.0
47		<1.0			

The data collected from the mammalian SpdSyn, being very similar to PfSpdSyn, was considered in further studies of the putrescine binding cavity.

2-Mercaptoethylamine has an IC_{50} against recombinant PfSpdSyn of 76 μM and a *in vitro* growth of specific cells of 254 μM. The binding mode of this inhibitor is not known and no additional data was available. MTA binds in the dcAdoMet cavity and is one of the products which forms during catalysis. MTA has an IC_{50} of 159 μM and was not tested for a *in vitro* growth of specific cells. AdoDATO, a transition-state analogue of the reacting substrates in SpdSyn is an inhibitor, which when bound to SpdSyn occupies the whole active site cavity and has been co-crystallized in SpdSyn of *T. maritima* (PDBid 1JQ3) and *P. falciparum* (PDBid 2I7C Korolev *et al.* (2002); Dufe *et al.* (2007)). Figure 3.12 illustrates the hydrogen bond network formation between PfSpdSyn (PDBid 2I7C) and AdoDATO. The hydrogen bond network of the adenosyl moiety of AdoDATO, MTA and dcAdoMet are very similar. The most important interactions involving hydrogen bonds can be observed between the adenosine ring and Asp 179, Pro 203 and Ala 179 (Figure 3.12). A hydrogen bond is also found with Ile 148 but is not always present in the PfSpdSyn crystal structures (Figure 3.12). The sugar ring of the adenosine moiety forms hydrogen bonds with Glu 147. The hydrogen bond formed between adenosine ring and Pro 203 is the only hydrogen bond formed between the moiety of interest and the gate-keeping loop and thus is thought to play a stabilizing role (Figure 3.12).

Knowledge obtained from a structure-based virtual screening study performed by Jacobsson *et al.* (2008) on PfSpdSyn was also included after its release in the second quarter of 2008. Interactions hypothesized to be important between AdoDATO and PfSpdSyn from the crystal structure, PDBid 2I7C were converted to PhFs (Jacobsson *et al.*, 2008). These PhFs were used to construct two pharmacophore models, one representing the dcAdoMet cavity and the second the putrescine binding cavity (Figure 3.13). The first consisted of six PhFs and can be broken down to the following: one hydrogen bond acceptor and one hydrogen bond donor corresponding to the backbone hydrogen bonds with Pro 203 and Ala 179, one aromatic ring corresponding to the adenine ring as well as two
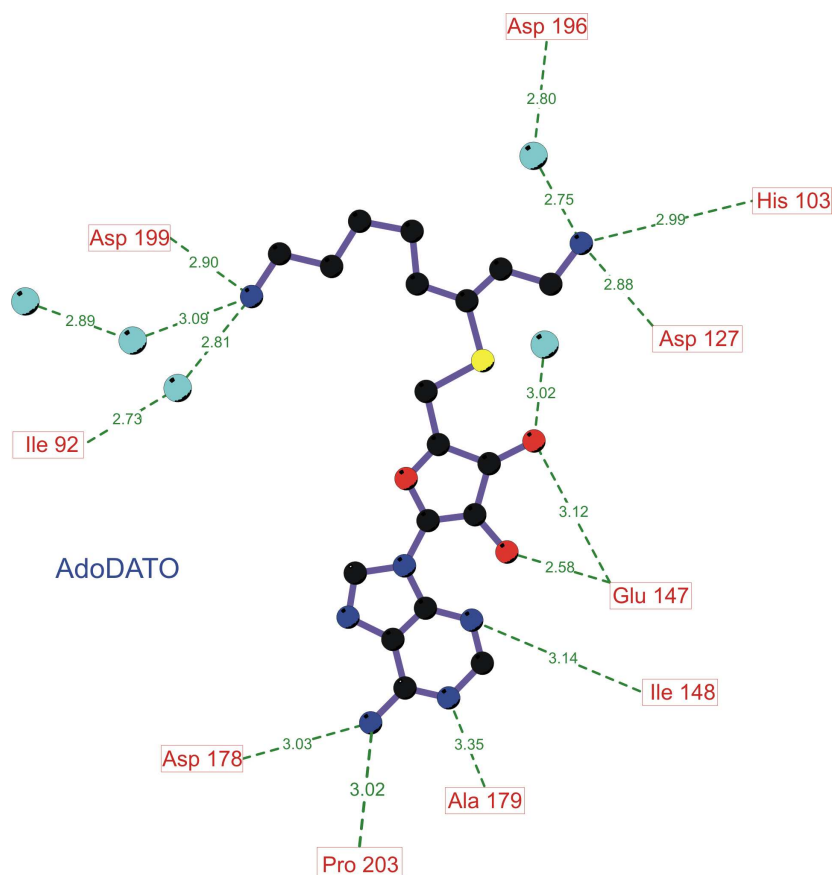
Figure 3.12: A 2D representation of the hydrogen bond network, which forms between PfSpdSyn (PDBid 2I7C) and AdoDATO. Interacting water molecules are represented by cyan spheres.

hydrogen bond donors and an acceptor from the sugar moiety of AdoDATO interacting with Glu 147 and Gln 72 (Figure 3.13). The pharmacophore model representing the putrescine binding cavity contained two PhFs, the first representative of a hydrophobic feature corresponding to the butyl moiety of putrescine and the other a positive ionizable group representing the non-attacking nitrogen of putrescine (Figure 3.13). These pharmacophore models were used in a six-step process to identify twenty eight compounds from a 2.6 million compound in-house chemical library and were subjected to saturation transfer difference (STD)-NMR and relaxation filtering experiments. From the 28 compounds seven were found to be reversible binders of PfSpdSyn and included compounds 4, 26, 1, 25, 17, 14 and 10 (Figure 3.14).

The six-step process used by Jacobsson *et al.* (2008) consisted of; 1) Phase flexible

a)   b)
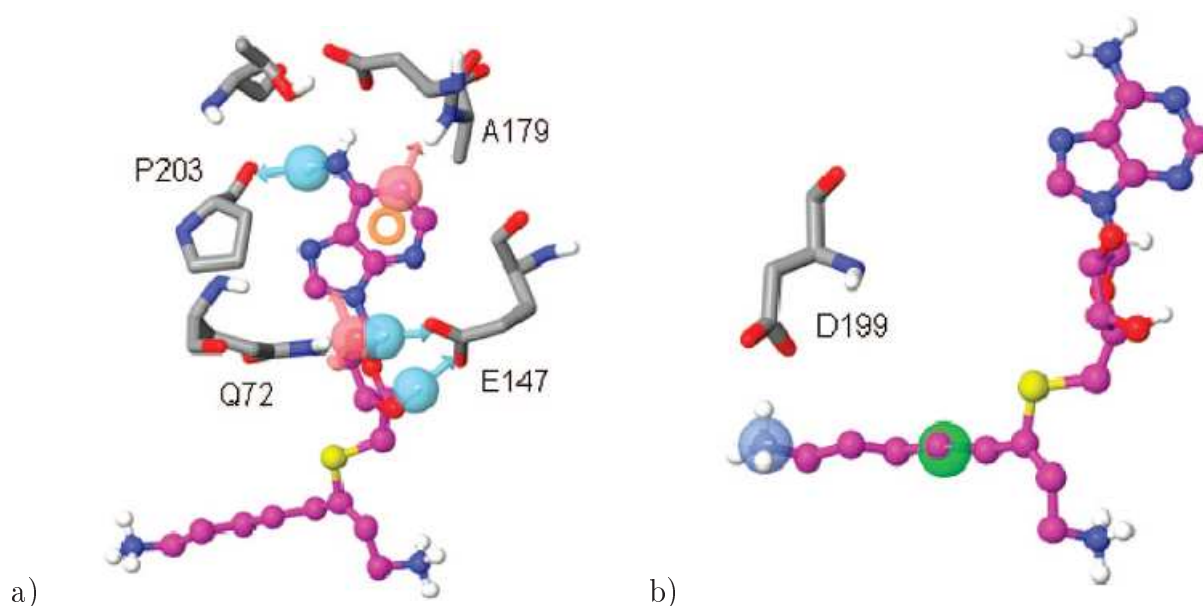
Figure 3.13: Phase pharmacophore models used in the virtual screen, representing interactions with the adenosine moiety of AdoDATO (a) and amine substrate moiety (putrescine binding cavity) of AdoDATO (b). AdoDATO is shown in ball and stick representation with carbons in magenta, selected residues from SpdSyn are shown in tube representation. Hydrogen bond donor features are shown in light red, acceptor features are shown in cyan, aromatic ring features are shown in orange, hydrophobic features are shown in green, and positive ionizable features are shown in blue (Adapted from Jacobsson *et al.* (2008)).

pharmacophore filtering, 2) Glide SP docking pose generation, 3) Phase rigid pharmacophore filtering, 4) Glide XP docking and scoring, 5) Visual inspection and 6) Experimental testing using STD-NMR and relaxation filtering experiments. The phase flexible pharmacophore filtering used means that the pharmacophore models were derived from the inhibitor AdoDATO and built from multi-conformers as to incorporate the flexibility of the ligand (Patel *et al.*, 2002). The compounds identified by the phase flexible pharmacophore filtering stage were docked using Glide SP with no constraints (Jacobsson *et al.*, 2008). The best nine docking poses of each compound were retained and fitted to the two pharmacophore models used. This allows the scientist to filter the compounds based on docking scores and fit values. This is known as the phase rigid pharmacophore filtering step and refers to the pharmacophore perception using only the X-ray ligand structures (Patel *et al.*, 2002). The best ranked compounds were then docked using the computationally more expensive docking program Glide XP (Jacobsson *et al.*, 2008; Patel *et al.*, 2002). From the latter docking study 193 compounds were selected as potential
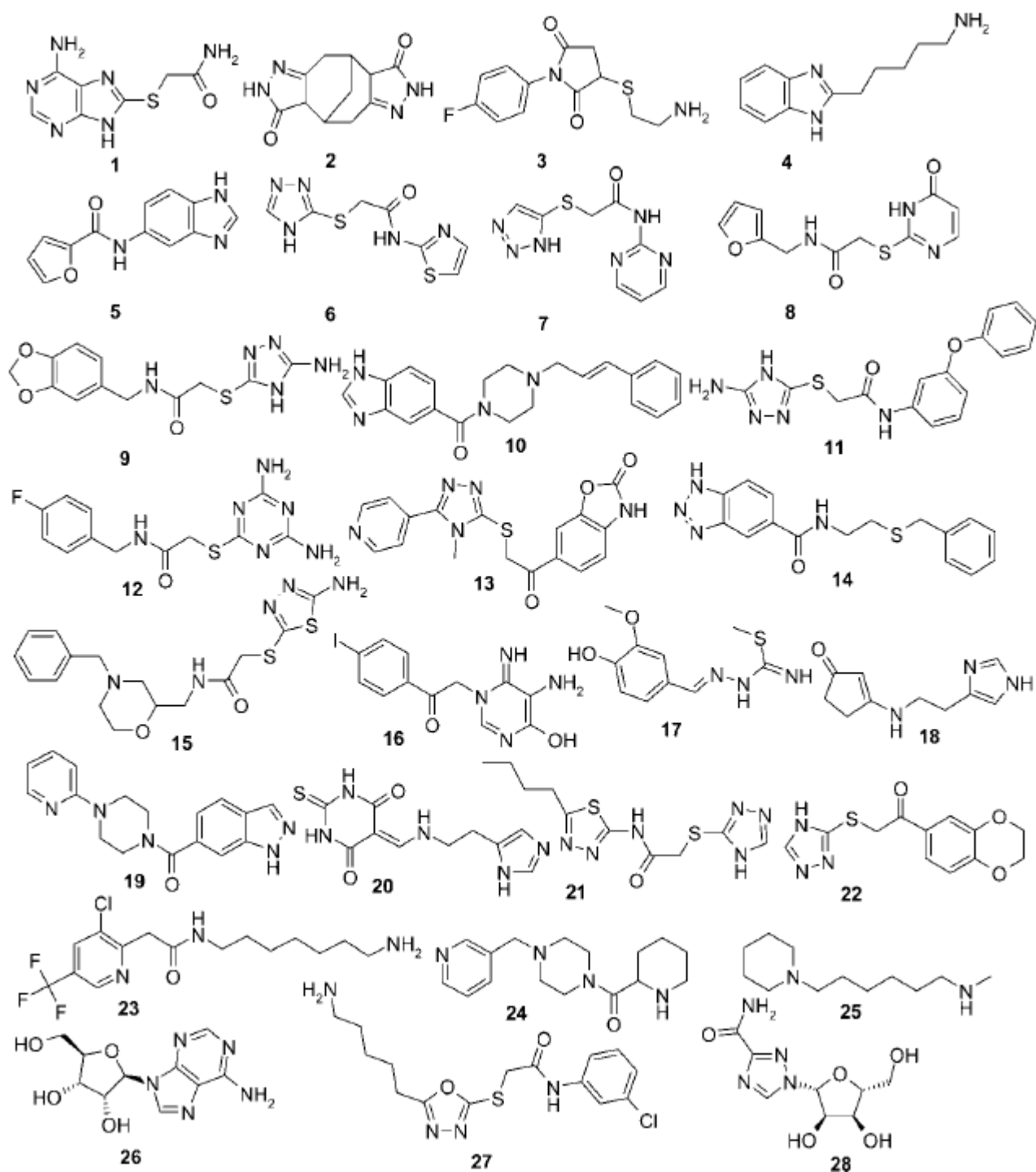
Figure 3.14: The 28 compounds identified by Jacobsson *et al.*, (2008) as potential PfSpdSyn active-site binders during their virtual screen (Adapted from Jacobsson *et al.* 2008).

inhibitors and then re-docked keeping the nine best docking poses. These poses were visually inspected and twenty eight compounds were selected to be tested using the STD-NMR method (Figure 3.14).

The study of Jacobsson *et al.* (2008) differs from the current study in that it used AdoDATO co-crystallized with PfSpdSyn to derive the pharmacophore models. The most important interactions between AdoDATO and PfSpdSyn were identified and used to describe its chemical properties using PhFs. These PhFs were then used to derive the pharmacophore models in a way similar to the analogue-based approach. In the current study the pharmacophore models used were derived from a negative image construction of the active site using various chemical entities to explore it. The current approach has the advantage that it is possible to identify novel PhFs, whereas the approach followed by Jacobsson *et al.* (2008) only allows for the use of PhFs, which can be derived from the interactions of AdoDATO. In the study of Jacobsson *et al.* (2008) only the flexibility of the ligands were incorporated during pharmacophore screening. The current study has the second advantage that it incorporates both the receptor's flexibility as well as the flexibility of the database ligands when screening the DPMs. This is possible since the flexibility of the receptor is captured within the DPMs. The flexibility of the database ligands are accounted for when screening a multi-conformer database with the BEST search algorithm of Catalyst which allows for the full optimization of the ligands during a database screen. The study of Jacobsson *et al.* (2008) relied heavily on docking of compounds to PfSpdSyn and in our experience docking to PfSpdSyn is not a trivial task. Also the approach followed in the current study is less automated and requires more decision making from the user giving more control over the development process.

3.2. Methods and Methodology

3.2.1. Protein Structure Quality Assessment and pKa Predictions

3.2.1.1. Protein Structure Quality Assessment

The PfSpdSyn structures, PDBid 2I7C and 2PT9, were prepared for further analysis by checking their structure integrity, using WHATIF (Vriend, 1990) and PROCHECK (Laskowski *et al.*, 1993). The structures were visually inspected and residues with alternative conformations were evaluated to find the conformations making the most chemical sense. The preferred residue conformation was selected based on the following criteria: 1) the residue conformation making the highest number of hydrogen bonds was selected and 2) hydrophobic residues conformations thought to be present in the most hydrophobic chemical space were selected. Surface residues were also evaluated using this criteria, however if no preference for the selection of either conformation could be justified, the conformation to be used was selected randomly.

3.2.1.2. pKa Prediction

The pKa predictions were performed using UHBD (`http://adrik.bchs.uh.edu/uhbd`) and YASARA (Krieger *et al.*, 2006). The calculations were performed for the PfSpdSyn crystal structures, PDBid 2HTE and 2I7C, and included crystal waters. pKa calculations were also performed for the PfSpdSyn structure (PDBid 2I7C) co-crystallized with AdoDATO using YASARA (Table 3.9). The pKa predictions performed using UHBD were done by adjusting the default scripts distributed with the package. The pKa values were predicted at pH 7. The temperature was set to 293K and the dielectric constant of the solvent and protein was kept at 80 and 20, respectively. The pKa values for both nitrogen atoms of histidine were predicted with UHBD. pKa predictions were not performed using UHBD in the presence of AdoDATO since no suitable parameters for AdoDATO could be implemented.

There are three protonation states that can be assigned to histidines: HSD, HSE and HSP. HSD is the protonation state where a neutral hydrogen is found on the ND1

atom, HSE where the neutral hydrogen is found on the NE2 atom and HSP where both ND1 and NE2 are protonated. UHBD can predict pKa values for both of the nitrogen atoms of histidine, however the nitrogen of interest needs to be assigned before either HISA or HISB. To predict the pKa of the ND1 atom one needs to select HISB, which consequently treats the NE2 atom in the neutral form. HISA in turn treats the ND1 atom in the neutral form and predicts the pKa of the NE2 atom. The pKa values of both the nitrogen atoms of histidine were predicted.

3.2.2. Phase Space Sampling

3.2.2.1. Molecular Dynamics

The MD simulations were performed in the absence of AdoDATO. All the energy minimizations and MD simulations were performed with NAMD (Phillips *et al.*, 2005) using the CHARMM force field (Brooks *et al.*, 1983). The solvation and neutralization of the system was performed using VMD (Humphrey *et al.*, 1996). The cell dimension of the water box was 79x100x81Å with the water boundary set 10 Å from the protein. In preparation for the PfSpdSyn equilibration run the protein was protonated according to the pKa studies discussed above. 100 steps of steepest descent (SD) minimization were performed on the added hydrogens followed by the solvation and neutralization of the protein. The water and ions in the water box were then minimized and equilibrated around the protein. This was done for 2 000 steps of SD minimization followed by 20ps equilibration of the water and ions with a reassignment of velocities every 1ps. The protein was then minimized for 200 steps using SD while keeping the solvent fixed. The protein and solvent were then heated to 310K. A heating gradient of 10K every 500 steps (1ps) was followed and the protein was kept at 310K for 34ps giving the process of heating a total duration time of 50ps. The protein was then equilibrated for 500ps and the temperature reassigned every 500 steps to 310K. Periodic boundary conditions (PBC) were applied and all the electrostatic interactions were included using the PME summation method. Constant pressure and temperature control were applied using the Berendsen method. The production run was performed for 5ns applying PBC. All the electrostatic interactions

were included using the PME summation. As with the equilibration run Berendsen dynamics were used to perform constant pressure and temperature control.

3.2.2.2. Clustering of Molecular Dynamics

The gromos algorithm (Xavier *et al.*, 1999) of the g_cluster module from GROMACS (Spoel *et al.*, 2005) was used to cluster the MD simulations and to obtain the central structure for the representative clusters. The clustering was done separately for each monomer of the simulated dimer. The MD trajectory was aligned and clustered, based on the active site of each of the monomers. The cutoff values used for the clustering were 1.15Å and 1.14Å for chain B and C, respectively.

3.2.3. Negative Image Construction

3.2.3.1. MIF Analysis

The program GRID (`http://www.moldiscovery.com`) was used for MIF analysis. A water probe was used to identify water binding hotspots within the active site of PfSpdSyn. The subensemble of structures selected in the previous section were prepared for MIF analysis. A grid box was generated to cover the active site with dimensions denoted in Table 3.5. Two water binding hotspots were identified and subsequently added to the subensemble of structures. MIF analysis was performed to explore the chemical nature of the active site by using probes representing HBDs, HBAs and HYD characteristics (PhFs). Three different HBD probes were used and included the N2 (the N2 probe represents a neutral flat nitrogen with two hydrogens (NH_2)), N1+ (the N1+ probe represents a sp^3 amine (NH) cation) and N3+ (the N3+ probe represents a sp^3 amine (NH_3 cation)) probes. For the identification of HBA binding hotspots three probes were used and included the O (the O probe represents a sp^2 carbonyl oxygen), O1 (the O1 probe represents an alkyl hydroxy (OH) group) and N:= (the N:= probe a sp2 nitrogen with lone pair) probes. The hydrophobic features were identified using both the DRY (the DRY probe is a general hydrophobic probe) and Me (the C3 probe represents a methyl CH_3 chemical moiety) probes. The default GRID parameters were used except for that of the NPLA (changed to 4) and ALMD (changed to 1) parameters.

Table 3.5: The 3D box dimensions used to cover the active site of the subensemble of PfSpdSyn structures. BOTX represents the X coordinate at the bottom of the grid box and the TOPX the top X coordinate.

BOTX	-19.82	TOPX	16.73
BOTY	-16.64	TOPY	20.85
BOTZ	-19.28	TOPZ	14.02

3.2.3.2. Pharmacophore Feature Identification

The binding energy value and coordinates of the grid points with the most favourable energies in the Z-plane were parsed from the GRID output file (*.lont) into a PDB file with the energy values represented in the Beta column using an in-house python script. This was done for all the protein structures of the subensemble. The newly generated PDB files were overlaid within the active site of PfSpdSyn. Clusters of grid points were identified and selected by visualizing the grid points within the active site. The pharmacophore features (PhF) were generated by extracting the attributes for the grid points from the identified clusters (grid point clusters) and calculating their center of mass (energy weighted) as well as their radius of gyration. This was performed on all the probes used in MIF analysis, using an in-house python script. This script generates a PDB file containing the center of mass (representing the PhF coordinates; energy weighted), the radius of gyration and the mean energy of the PhF. The PhFs identified for the subensemble of PfSpdSyn and those from the static PfSpdSyn crystal structure (PDBid 2PT9) were stored separately.

3.2.3.3. Exclusion volumes

Atoms to represent the EVs for the DPMs were selected after a visual inspection of residues surrounding the active site. The coordinates for these atoms were extracted from all the structures in the subensemble of PfSpdSyn. The center of mass (mean of the coordinates) for these atoms was calculated to include the dynamic behavior of the active site and thus represent "dynamic" EVs. This was also performed using an in-house python script.

3.2.3.4. Pharmacophore Model Selection

For the selection of DPMs the active site was subdivided into four regions, DPM1 through to DPM4. For each of these regions various DPMs represented by different combinations of PhFs were constructed. The DPMs were built in Catalyst v4.10. The EVs were added using Discovery Studio 2.0 before screening subsets of the ZINC database (`http://www.accelrys.com`).

3.2.4. Hit Analysis and *In Vitro* Testing

3.2.4.1. Database Construction and Searching

The drug-like subset of the ZINC database was used to construct a multiconformer composite database. These subsets were first screened for duplicates, which were removed using an in-house python script. The drug-like subset contained 2 011 000 unique entries. The multiconformer composite database was generated using catDB from Accelrys (`http://www.accelrys.com`). During database construction the maximum conformers were limited to 250.

Catalyst v4.10 was used to search both the constructed databases selecting the BEST flexible search parameter. The DPMs used during searching included the EVs generated from the subensemble of PfSpdSyn. Compounds identified during these searches were fitted to their respective DPMs to get the best fitting compounds and were ranked accordingly. The EVs were present during the fitting of the compounds to the DPMs. Visual inspection of the best fitting compounds was performed to select the top compounds based on their fit values and orientation within the active site.

3.2.4.2. Docking and Compound Selection

Compounds identified during the database searches were docked using AutoDock 4 (Morris *et al.*, 1998). Two monomers of different PfSpdSyn crystal structures were prepared for docking and included chain C of PfSpdSyn, PDBid 2I7C and chain A of PfSpdSyn, PDBid 2PT9. Both these structures were prepared for docking by subjecting them to 400 steps of steepest descent energy minimization. The AutoDockTools (ADT)

kit was used in further preparation of both the target structures and the compounds identified during database searching. The parameters changed for docking were as follows, the Genetic Search Algorithm was used and the number of genetic algorithm (GA) runs were changed from 10 to 50, the Translation parameter was changed from 2.0 to 0.2 (Angstrom/step) whereas both the Quaternion and Torsion parameters were changed from 50 to 5 (Degree/step). The RMS Cluster tolerance (Angstrom) was changed from 2.0 to 1.5. The docked compounds were evaluated based on their energy scores and poses within the active site of both the PfSpdSyn target structures used during docking.

3.3. Results and Discussion

3.3.1. Protein Structure Quality Assessment and pKa Predictions

3.3.1.1. Protein Selection and Structure Quality Assessment

Seven crystal structures are currently available for PfSpdSyn in the PDB of which six were crystallized in complex with ligands (Table 3.6). PfSpdSyn (PDBid 2I7C) was crystallized in complex with AdoDATO and was the first PfSpdSyn structure to be released with both its monomer's gate-keeping loops resolved. The structure was shown to have a resolution of 1.7Å and can therefore be considered as a high resolution structure. The R-factor and R_{free}-factor values were shown to be 0.176 and 0.194 respectively, emphasizing the quality of the crystal structure. It was subsequently selected to be used in the development of a DPM for PfSpdSyn (Table 3.6).

Table 3.6: Available crystal structures of PfSpdSyn (Oct 2008)

PDBid	Release date	Resolution (Å)	R-factor	R_{free}-factor	Ligand
2HTE	08-Aug-2006	2.00	0.188	0.231	MTA
2I7C	12-Sep-2006	1.71	0.176	0.194	AdoDATO
2PWP	22-May-2007	2.10	0.195	0.232	spermidine
3B7P	20-Nov-2007	2.00	0.216	0.250	spermine
2PT9	01-Apr-2008	2.20	0.198	0.239	dcAdoMet and 4MCHA
2PT6	01-Apr-2008	2.00	0.182	0.214	dcAdoMet
2PSS	01-Apr-2008	2.20	0.207	0.247	apo-form

The PfSpdSyn dimer (chains B and C) was subjected to stereo-chemical analysis using PROCHECK and WHATIF and a Ramachandran plot was generated (Figure 3.15). Chain B has 89.3% of its residues in the most favoured region and the other 10.7% in the additional allowed region. Chain C has 88.1% in the most favoured region, 11.5% in the additional allowed region and 0.4% in the generously allowed region. The stereochemical checks performed using both WHATIF and PROCHECK were found to be in the generally acceptable ranges and included the Chi1-Chi2 plots, main-chain and side-chain parameters, main-chain bond lengths and angles, RMS distances for planarity and distorted geometry. Visual inspection was used to choose the most relevant conformations of amino acid with double entries and denoted in Table 3.7. The selection criteria was described in the methods section 3.2.
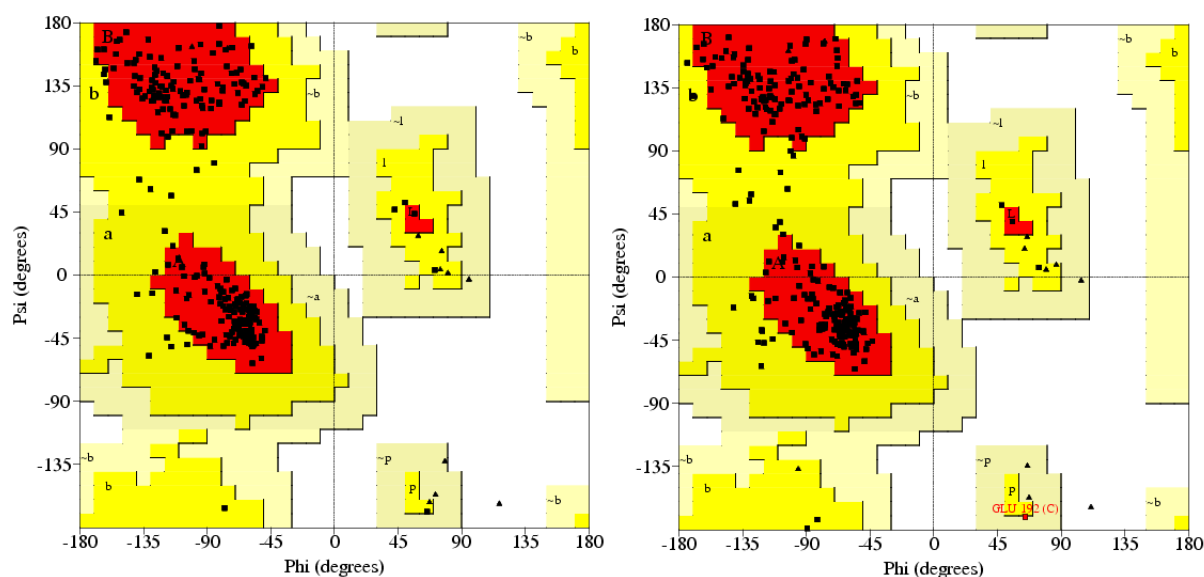
Figure 3.15: Ramachandran plots for the PfSpdSyn crystal structure dimer, PDBid 2I7C. The Ramachandran plot on the *left* constitutes chain B and the plot on the *right* chain C. Regions in red correspond to conformations where there are no steric clashes whereas yellow regions represent regions where slightly shorter van der Waals radi were used in calculation.

The crystal structure of PfSpdSyn resolved in complex with dcAdoMet and 4MCHA (PDBid 2PT9) was released to us earlier and used in this study. The coordinates were provided by Salam Al-Karadaghi, affiliated with the Department of Molecular Biophysics, Center for Molecular Protein Science, Lund University, Sweden. The same approach of

Table 3.7:   The side chain conformations selected for residues of PfSpdSyn crystal structure (PDBid 2I7C) having more than one conformation.

PfSpdSyn PDBid 2I7C		
Residue	Conformer Chosen	
	chain B	chain C
Met 50	B	A
Ser 156	B	A
Asn 185	A	N/A
Ser 197	B	N/A
Ile 216	A	N/A
Asn 218	A	N/A
Ser 305	B	B
PfSpdSyn PDBid 2PT9		
Glu 154	N/A	A
Val 156	N/A	A

N/A Not Applicable

assessing the quality of the PfSpdSyn structure of PDBid 2I7C was followed. The crystal structure (PDBid 2PT9) has a resolution of 2.20Å, which is within the acceptable range of a high quality structure. The R-factor and R_{free}-factor were found to be 0.198 and 0.239 also supporting the good quality of the crystal structure. The stereochemical analysis on the dimer showed that chain B has 90.1% of its residues in the most favoured region, 9.5% of the residues in the additional allowed region and 0.4% in the disallowed region, which were relieved after energy minimization (Figure 3.16). Chain C showed that 88.1% of its residues are in the most favourable region, 11.5% in the additional allowed region and 0.4% in the generously allowed region (Figure 3.16). The stereochemical checks performed using both WHATIF and PROCHECK were found to be in the generally acceptable ranges as with PfSpdSyn (PDBid 2I7C). From the stereo-chemical analysis it can be concluded that the structure is of high quality. Visual inspection was performed on the structure and only two residues with alternative conformations were found and selected as descibed in section 3.2 (Table 3.7).

3.3.1.2. pKa Prediction

A protein consists of various titratable groups which can assume various protonation states and these need to be assigned correctly *a priori* for molecular dynamics. Prior to pKa predictions a visual inspection of the PfSpdSyn crystal structures (PDBid 2I7C
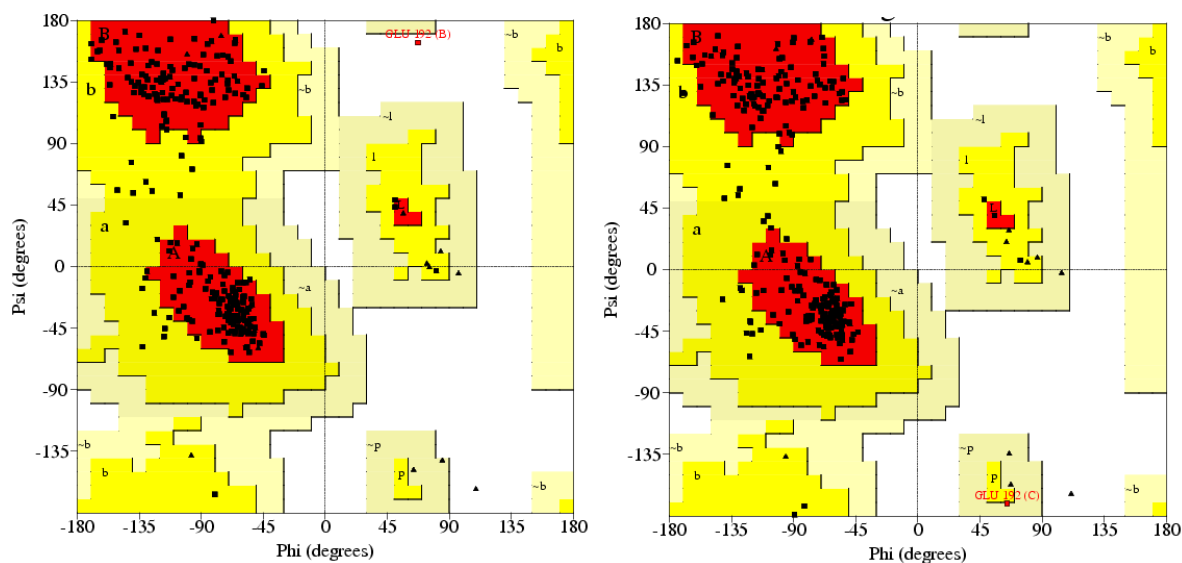
Figure 3.16: Ramachandran plot for the dimer of the PfSpdSyn crystal structure, PDBid 2PT9. The Ramachandran plot on the *left* represents chain B and the one on the *right* chain C.

and 2HTE) was performed to guide the assignment of protonation states, especially with regards to the histidine residues. The nitrogen atoms of the respective histidines were evaluated and HBDs and HBAs within hydrogen bond formation distances were identified and denoted (Table 3.8). The information was used to guide pKa predictions and the assignment of the protonation states of the respective histidine atoms. Four histidines occur in each monomer of PfSpdSyn, His 103, His 108, His 236 and His 304. The predicted pKa values of the histidines for both the monomers of interest are tabulated in Table 3.9. Histidine 103 was the only His to be predicted in the HSD state and the rest were predicted to be in the HSE state.

Visual inspection of PfSpdSyn revealed that the ND1 nitrogen atom of His 103, can potentially make a hydrogen bond with the carboxylic group of Glu 99 in both chains B and C (Figure 3.17(a)). It is also possible for the NE2 nitrogen atom to make a hydrogen bond with the aminopropyl group of AdoDATO (Figure 3.17(a)). A pKa prediction of both the nitrogens was made (Table 3.9). Predictions with only the crystal waters of the PfSpdSyn crystal structure (PDBid 2I7C) present using UHBD, revealed pKa values of 8.15 (HISA; NE2), 7.98 (HISB; ND1) and 8.96 (HISA; NE2), 9.12 (HISB; ND1) for chain B and C, respectively (Table 3.9). The YASARA program was also used to predict

Table 3.8: Histidine residues of PfSpdSyn (PDBid 2I7C) and the proposed interaction with the ND1 and NE2 atoms obtained from visual inspection.

Atom Num	HIS assigned	HIS atom	Interacting atom	Interacting residue	Distance
Chain B					
103	HSD[▷]	ND1	OE2	GLU 99	2.79
		NE2	Binds aminopropyl moiety of AdoDATO		
108	HSE[◁]	ND1	OH	TYR 137	2.77
		NE2	OE1	GLU 133	2.81
236	HSE[◁]	ND1	N	SER232	2.90
		NE2	O	MET 50	3.13
304	HSE[◁]	ND1	OH	TYR 299	2.90
		NE2	O	ILE 281	2.81
Chain C					
103	HSD[▷]	ND1	OE1	GLU 99	2.76
		NE2	Binds aminopropyl moiety of AdoDATO		
108	HSE[◁]	ND1	OH	TYR 137	2.76
		NE2	OE1	GLU 133	2.81
236	HSE[◁]	ND1	N	SER232	3.04
		NE2	O	MET 50	2.87
304	HSE[◁]	ND1	OH	TYR 299	2.71
		NE2	O	ILE 281	2.77

[▷] HSD means that His is in a neutral state and that the ND atom is protonated.

[◁] HSE means that His is in the neutral state and that the NE atom is protonated.

the pKa values of the crystal structure (PDBid 2I7C) containing only crystal waters and found to be 8.04 and 8.92 for chains B and C, respectively. Both these programs predicted His 103 to be protonated. However, when considering the PfSpdSyn crystal structure, PDBid 2HTE (chain C) containing MTA, it is noted that a water molecule was crystallized within 2.73Å from the NE2 atom of His 103 (Figure 3.18). This implies that it is possible for the NE2 nitrogen of His 103 to make a hydrogen bond with this water. pKa predictions performed on PfSpdSyn (PDBid 2HTE (chain C)) containing the crystal waters and MTA with YASARA revealed a value of 8.12. It is however believed that His 103 is in the HSD state and not protonated (HSP), since His 103 plays an important role in the binding of the aminopropyl moiety of dcAdoMet, which is protonated and hence, it therefore does not make chemical sense that His 103 would be in the protonated form. Therefore, the His103 was assigned as HSD. His 108, 236 and 304 were assigned to be in the HSE state, based on the chemical knowledge gained from visual inspection and
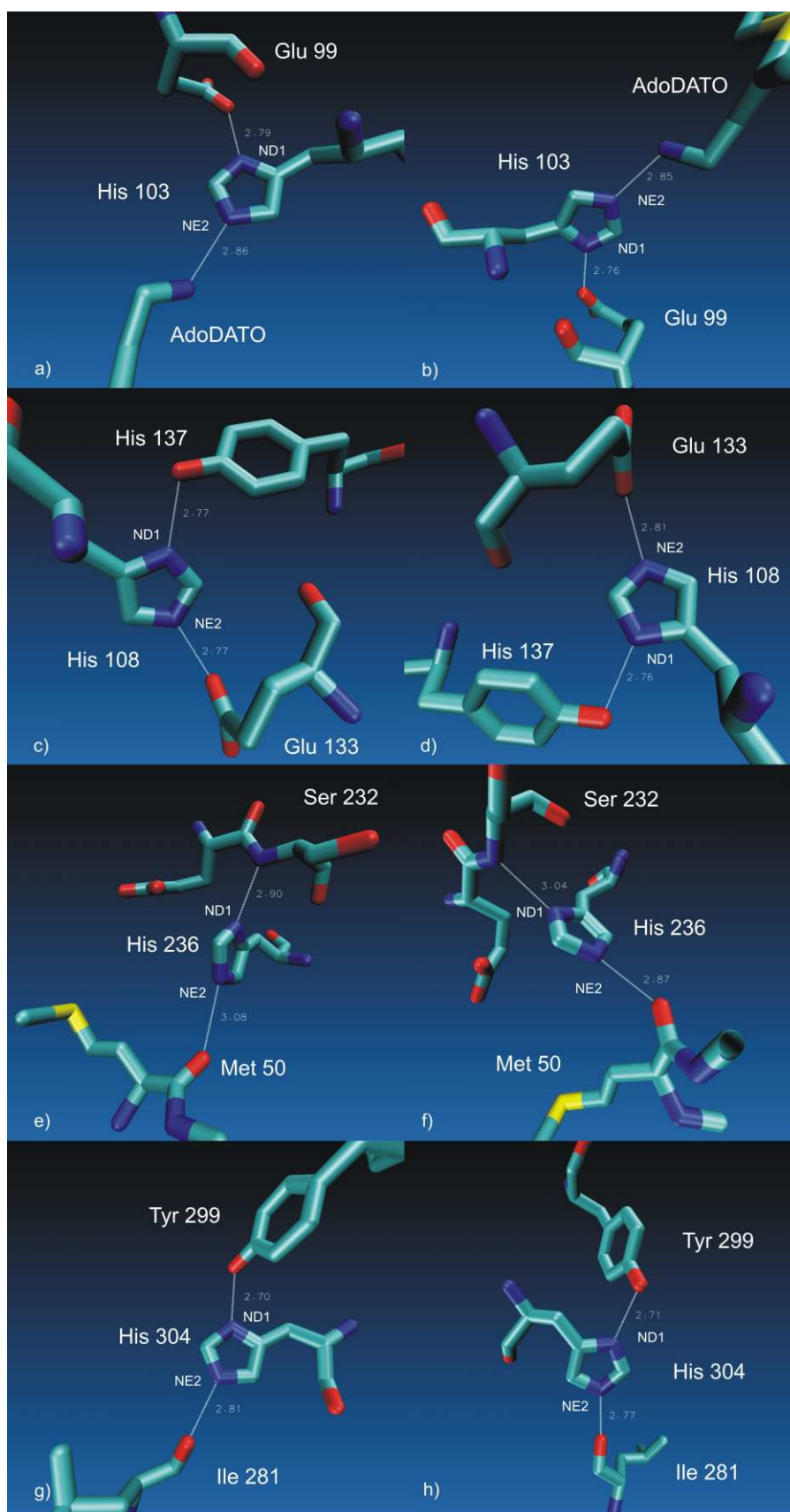
Figure 3.17: A visual representation of the immediate environment of histidines present in both the monomers of the PfSpdSyn crystal structure, PDBid 2I7C (chain B *(left)* and chain C *(right)*). The histidine's are represented as follows: His 103 (**a)** *(chain B)* - **b)** *(chain C)*), His 108 (**c)** *(chain B)* - **d)** *(chain C)*), His 236 (**e)** *(chain B)* - **f)** *(chain C)*) and His 304 (**g)** *(chain B)* - **h)** *(chain C)*).

Table 3.9:   pKa predictions for the His residues of PfSpdSyn (PDBid 2I7C) calculated using UHBD and YASARA packages.

Program	UHBD						YASARA					HIS type
Chain	B		C		2HTEC		B	C	B*	C*	2HTEC	assigned
Residue	HISA‡	HISB‡	HISA‡	HISB‡	HISA‡	HISB‡	HIS	HIS	HIS	HIS	HIS	
103	8.15	7.98	8.96	9.12	8.51	9.29	8.04	8.92	6.25	6.34	8.19	HSD$^\triangleright$
108	7.13	6.24	7.53	6.74	7.27	6.33	7.51	7.63	7.20	6.62	7.65	HSE$^\triangleleft$
236	6.62	5.99	7.68	7.05	6.60	5.85	8.21	8.25	7.33	7.24	8.28	HSE$^\triangleleft$
304	5.85	5.65	6.47	6.38	6.03	5.84	6.81	7.00	6.55	6.70	8.28	HSE$^\triangleleft$

*pKa prediction for the PfSpdSyn crystal structure (2I7C) containing AdoDATO.

‡ HISA means that during pKa calculations the ND atom of HIS is treated as being in the neutral form, whereas HISB means that during the NE atom of HIS is treated in the neutral form.

$^\triangleright$HSD means that His is in a neutral state and that the ND atom is protonated.

$^\triangleleft$HSE means that His is in the neutral state and that the NE atom is protonated.

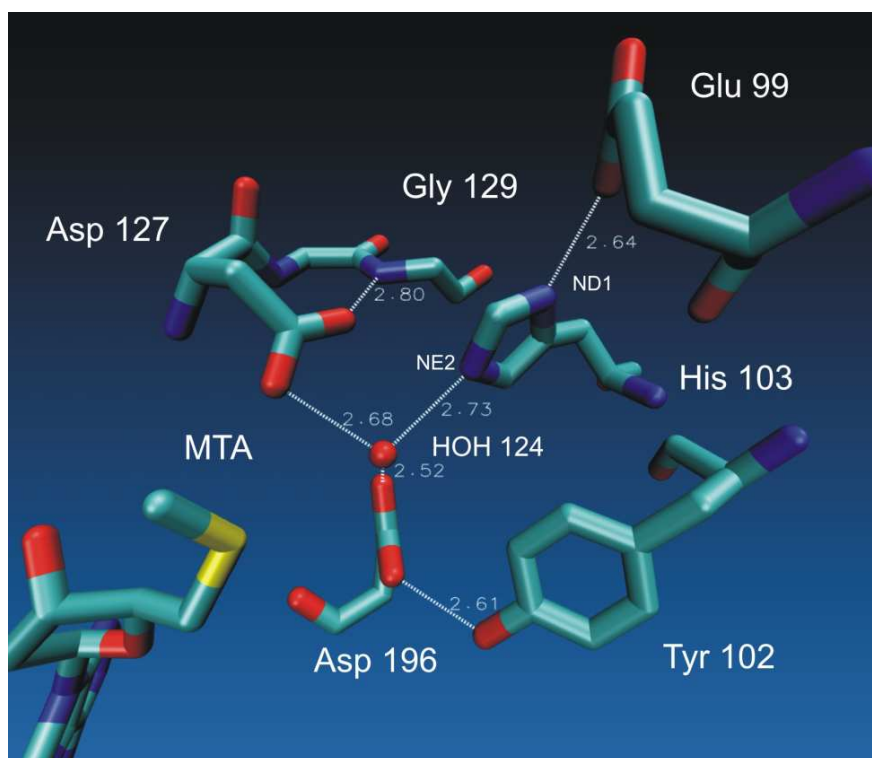the results obtained from the predictive programs (Figure 3.17(b), (c) and (d) and Table 3.9).

Figure 3.18: A visual representation of His103 from the PfSpdSyn crystal structure, PDBid 2HTE (Chain C).

3.3.2. Phase Space Sampling

3.3.2.1. Molecular Dynamics

The development of a dynamic receptor-based pharmacophore is based on using a subensemble of protein structures representing the dynamic behavior of the protein. Molecular dynamics simulations are used to estimate the equilibrium and dynamic properties of a complex. In the current study it was used to sample conformational space and to find the most represented structures within the sampled space. All the MD simulations were performed in the absence of AdoDATO, in order to obtain the representative state of the PfSpdSyn active site without biasing from interactions with ligands.

From the RMS coordinate deviation of the equilibration run it could be seen that the protein was converging, meaning the fluctuations occur around the mean for both chains B and C (Figure 3.19). The total energy of the equilibration run was also shown to be stable (Figure 3.20). These results gave confidence that the system was stable and subsequently a production run was performed.
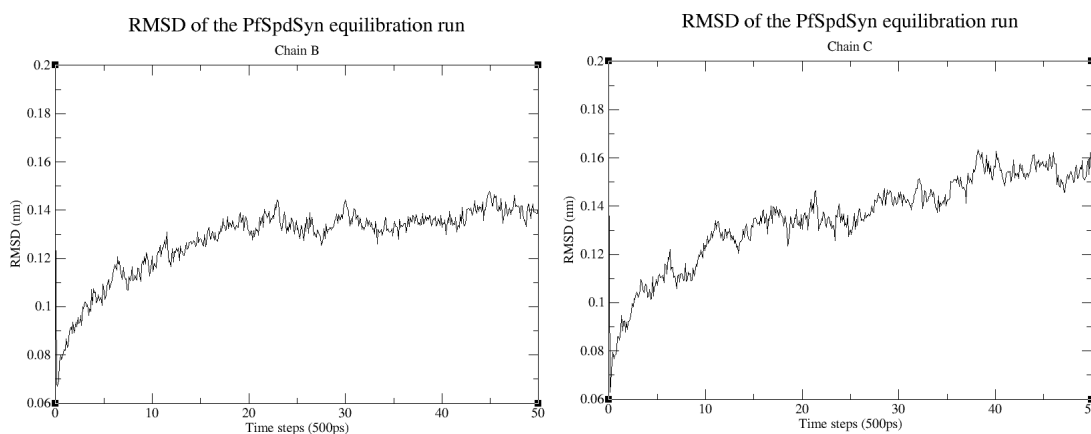

Figure 3.19: The RMS coordinate deviation for both chain B *(left)* and chain C *(right)* of the PfSpdSyn equilibration run.

The production run was performed for 5ns using PBC with all electrostatic interactions included by applying the PME summation method. The RMS coordinate deviation of the production run showed that the first 2.5ns were similar for both chains B and
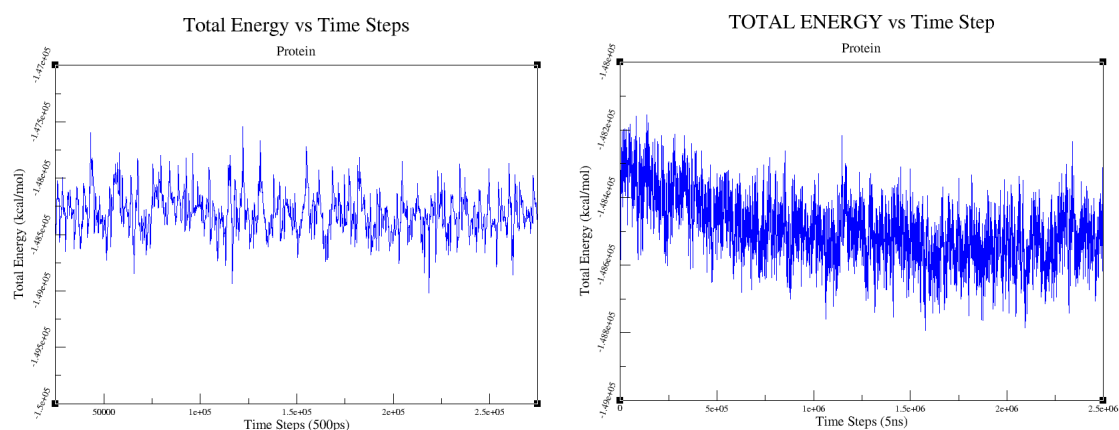
Figure 3.20: The total energy of PfSpdSyn during both the equilibration *(left)* and production *(right)* runs.

C (Figure 3.21). The RMS coordinate deviation of chain C was shown to increase after about 2.5ns and can be explained by the increase of loop movement from the gate-keeping loop. However, this was not found to be the case with chain B which indicated that the RMSD converges throughout the entire simulation. The total energy for the production run was plotted over time and found to be decreasing marginally (Figure 3.20). The results from the molecular dynamics were in accordance to what was expected of a stable system and were therefore used in further analysis.

Figure 3.21: The RMS coordinate deviation for both chain B *(left)* and chain C *(right)* of the PfSpdSyn production run.

3.3.2.2. Clustering of Molecular Dynamics Trajectory

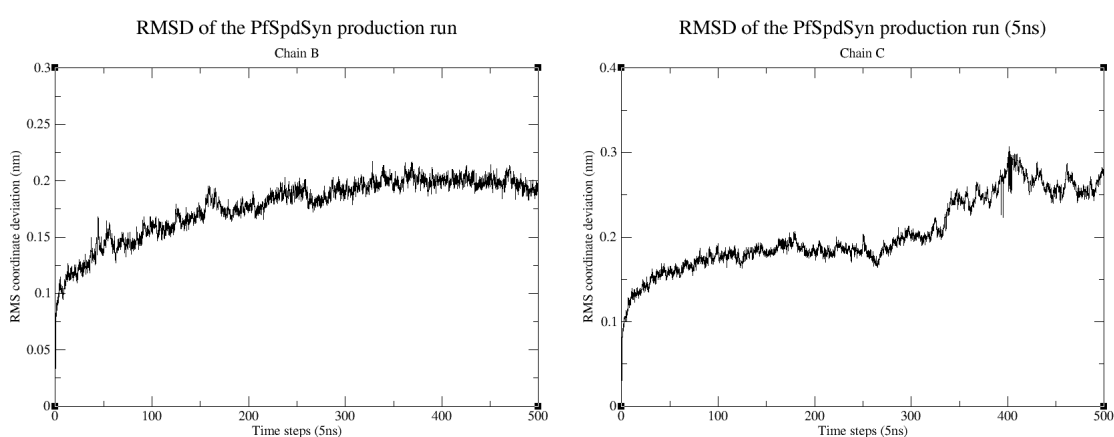The clustering of the MD production run for PfSpdSyn is a means to find the most representative structures of the sampled phase space. A selection of residues within 7Å of AdoDATO in the PfSpdSyn crystal structure (PDBid 2I7C, chain C) was made to get a representative group of residues for the active site. These selected residues were used in the clustering of the trajectory rather than the backbone atoms of PfSpdSyn. This ensured a better sampling of the conformational changes which occur within the active site during the MD simulation. Fifty two residues were selected to be part of this subset of atoms and are listed in Table 3.10.

Table 3.10: The active site residues selected and used for the determination of RMS coordinate deviation during the MD simulation.

Active site Residues			
Glu 46	Tyr 102	Val 152	Ala 204
Ser 48	His 103	Glu 177	Glu 205
Ile 49	Val 122	Asp 178	Thr 206
Met 50	Val 123	Ala 179	Lue 207
Trp 51	Gly 124	Ser 180	Phe 208
Tyr 71	Gly 125	Asp 196	Gln 229
Gln 72	Gly 126	Ser 197	Glu 231
Leu 88	Asp 127	Ser 198	Ile 235
Asp 89	Ile 130	Asp 199	His 236
Val 91	Cys 146	Pro 200	Thr 263
Ile 92	Glu 147	Ile 201	Tyr 264
Gln 93	Ile 148	Gly 202	Pro 265
Leu 94	Asp 149	Pro 203	Ile 269

The hierarchical gromos (Xavier *et al.*, 1999) algorithm of GROMACS was used to cluster the MD trajectory. This algorithm takes as input parameter a cutoff value, which can be varied to obtain the desired representative clusters. This means that various cutoff values need to be used to identify five clusters representing ∼90% of the structures captured during the MD simulation. The different monomers were treated separately, which in theory doubles the sampled space, since it is now possible to compare PfSpdSyn from two different starting structures (chains B and C), each simulated for 5ns.
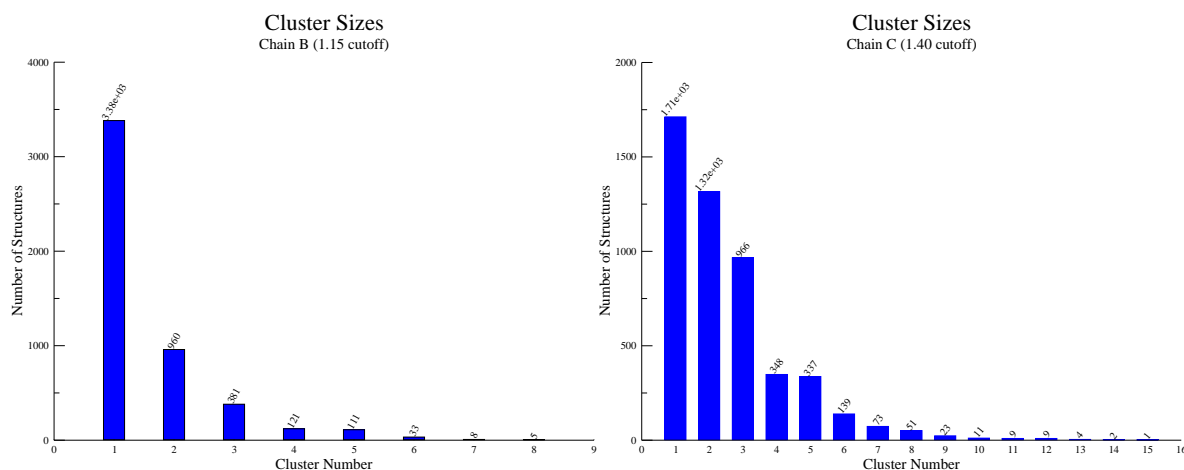
Figure 3.22: Clustering results of the PfSpdSyn production run for both the monomers. A cutoff value of 1.15Å was used for chain B *(left)* and a cutoff value of 1.40Å for chain C *(right)*. The bar graphs represent the number of structures in each cluster.

Chain B was clustered with a cutoff value of 0.115 (1.15Å) resulting in eight clusters having most of the structures in the five largest clusters (99.1%; Figure 3.22). Chain C was clustered with a cutoff value of 0.140 (1.40Å). Fifteen clusters were generated with 94.3% of the structures represented in the five largest clusters (Figure 3.22). The cutoff value for chain C was expected to be larger since some loop movement was observed in the last part of the MD simulation. For both chains the representative structures of the five largest clusters were overlaid with the active site residues displayed to illustrate the phase space sampled (Figure 3.23).

Table 3.11 denotes the RMSD values for the representative structures of the top five clusters obtained after clustering with GROMACS. The structures were aligned based on their backbone atoms before calculating the RMSD of the active site to determine the most divergent conformations (this was done within the VMD package). GROMACS does not allow one to first align the trajectory structures and then cluster based on the RMSD values of the active site, this being the reason for not clustering the MD trajectory using such a strategy. The RMSD analysis of the top ten clusters showed a RMSD range from 0.000 to 2.235 when aligned to the structure of cluster 1, chain B (Clus1B; Table 3.11). From the ten cluster structures, six were selected covering the whole RMSD range, equally distributed. The percentile of structures represented by each cluster were
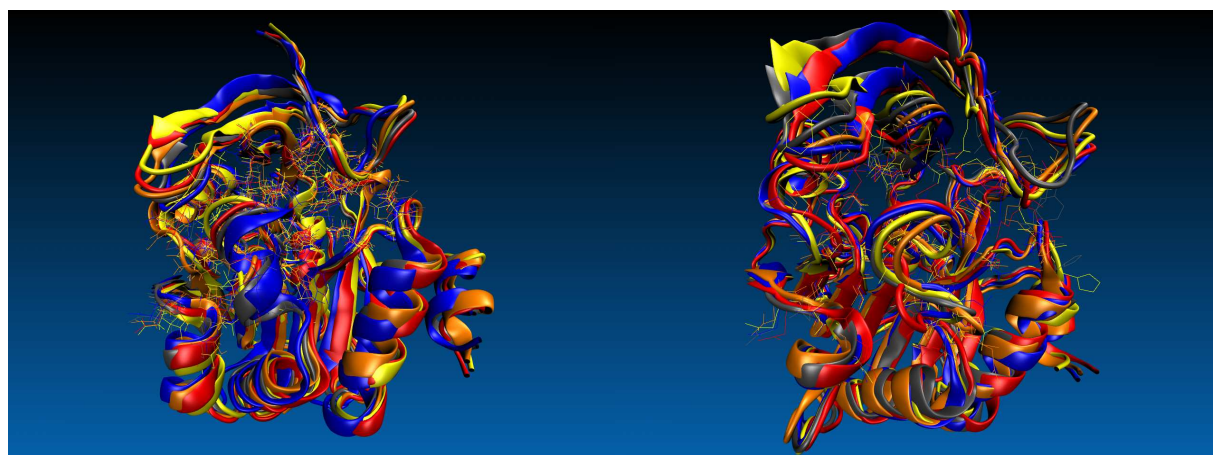
Figure 3.23: The representative PfSpdSyn structures of the five largest clusters overlaid (cartoon) for both monomers, chain B *(left)* and chain C *(right)*. Cluster 1 is represented in blue, cluster 2 in red, cluster 3 in gray, cluster 4 in orange and cluster 5 in yellow. The active site residues of the representative clusters are represented as lines.

considered when selecting the six most representative structures. For example Cluster 1 from chain C (Clus1C) was selected over Cluster 2 (chain C; Clus2C), Cluster 3 (chain C; Clus3C) and Cluster 4 (chain B; Clus4B) falling within the same RMSD range (RMSD range between 1.968 and 2.005; Table 3.11), the reason being that Cluster 1 (chain C) represent the highest percentage of structure (34.24%) from the MD trajectory and is thought to be the most representative. The six selected structures were (Table 3.11): 1) Clus1B, 2) Clus5B, 3) Clus3B, 4) Clus2B, 5) Clus1C and 6) Clus5C. These six selected structures thus constitute the subensemble of structures having captured within them the flexibility of PfSpdSyn and were subsequently used in further analysis to develop a DPM.

A comparison between the starting structure used for the MD simulations and the six structures identified to represent the subensemble of structures was made. The subensemble of structures was also compared to the three monomers resolved in complex with dcAdoMet and 4MCHA (PDBid 2PT9). The comparison with PfSpdSyn (PDBid 2PT9) was made to identifiy changes the active site undergoes upon binding of 4MCHA as to incorporate this knowledge at later stages of this study. It was found that Gln 229 undergoes large conformational changes compared to the starting structure. During the

Table 3.11: The RMS coordinate deviation values of the representative structures of the five biggest clusters of both the PfSpdSyn monomers, which were used to select the best representative structures for further analysis. The selected structures are indicated in the Selected Structures column labeled from one to six.

Protein	Cα Backbone RMSD	Active Site RMSD	% Representation	Selected Structures
		Chain B		
Clus1B	0.000	0.000	67.64	1
Clus2B	1.397	1.693	19.20	4
Clus3B	1.128	1.170	6.62	3
Clus4B	1.448	1.968	0.24	—
Clus5B	0.679	0.664	0.22	2
		Chain C		
Clus1C	1.591	1.976	34.24	5
Clus2C	1.556	1.990	26.32	—
Clus3C	1.590	2.005	19.32	—
Clus4C	1.381	1.737	6.96	—
Clus5C	1.747	2.235	6.74	6

MD simulation Gln 229 adopts a conformation which is similar to its conformation found in the apo (without ligand) structure of PfSpdSyn (PDBid 2PSS). This conformational change of Gln 229 alters the binding characteristics of the active site quite drastically. The conformational change Gln 229 undergoes upon binding of 4MCHA can be seen in Figure 3.24 *(Left)* and will be described using the bond between its CD and CG carbons (Figure 3.24). The CD-CG bond is perpendicular to the binding plane of 4MCHA in the subensemble of structures, however upon binding of 4MCHA it assumes a position which is parallel to the binding plane (Figure 3.24*(Left)*). It can be seen from Figure 3.24*(Left)* that the conformation Gln 229 adopts during the MD simulation will not allow for the binding of 4MCHA, since the CD-CG bond is perpendicular to the binding plane of 4MCHA. This conformation presents the amide group of Gln 229 to the binding cavity resulting in both steric and electrostatic clashes with 4MCHA.

Upon binding of 4MCHA the amide group of Gln 229 forms hydrogen bonds with Ser 198 and Glu 231 (with both the carboxylic group and backbone nitrogen), stabilizing this conformation (interactions not shown). It should also be noted that the CD carbon of Gln 229 is presented to the active site as to form hydrophobic interactions with 4MCHA (Figure 3.24 *(Left)*). For PfSpdSyn this conformational change was also observed in the
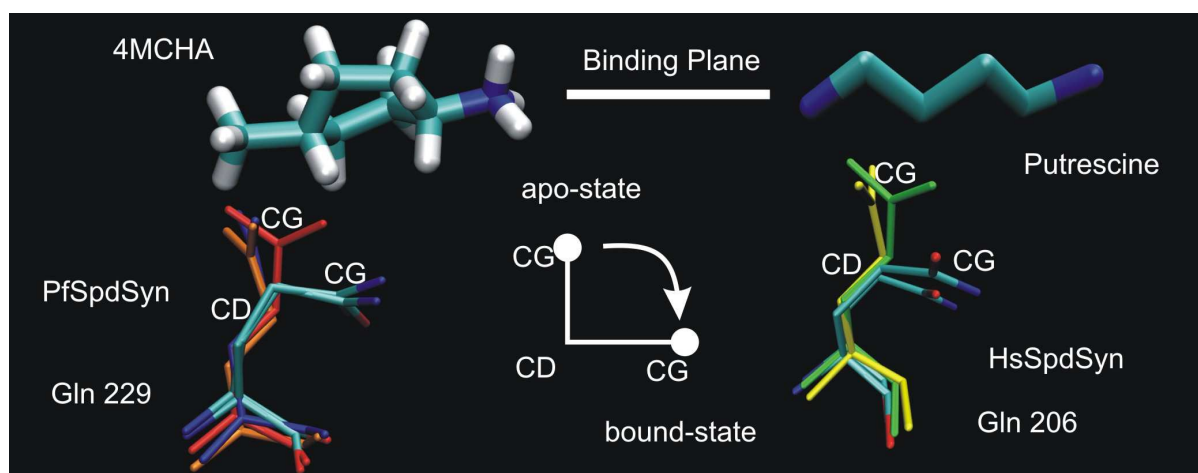
Figure 3.24: The conformational change Gln 229 undergoes upon binding of the ligands 4MCHA and AdoDATO as well as the naturally occurring substrate putrescine. For reasons of simplicity not all the residues from the clustered structures nor all the structures themselves are displayed. A schematic representation of the conformational change that occurs between the apo and bound-state for Gln 229 is displayed in white. It is explained using the orientation of the CD-CG bond of Gln 229 with reference to an assigned binding plane for 4MCHA and putrescine (white). *Left:* 4MCHA and Gln 229 of the crystal structure (PDBid 2PT9, chain A and B) are shown in a cyan color scheme. The blue, red and orange residues represents Clus1B, Clus2B and Clus5B, respectively and illustrates the different conformations Gln 229 adopts during the MD simulation (apo-state). *Right:* Putrescine co-crystallized with the human SpdSyn (PDBid 2O06) and the conformational change of Gln 206 which corresponds to Gln 229 of PfSpdSyn, represented in the cyan color scheme. The Gln 229 residues of PfSpdSyn from Clus1B and Clus2B are represented in yellow and green respectively and are shown to highlight the differences in orientation of Gln 206 within the active site of both the apo and bound states of SpdSyn.

structures resolved in complex with AdoDATO, spermidine and spermine. This conformational change was also found upon binding of putrescine to the HsSpdSyn (Figure 3.24 *(Right)*).

It is clear from Figure 3.24 that the CD carbon of Gln 229 presents HYD characteristics to the active site which interacts with ligands and substrates upon binding. Within the same chemical space the subsensemble of stuctures presents the amide group of Gln 229 having both HBDs and HBAs characteristics. These chemical differences will significantly change PhF identification during the negative image construction stage. Therefore, althought the conformation Gln 229 adopted during the MD simulation was confirmed by the apo-PfSpdSyn (PDBid 2PSS) structure, it is thought that using only the subsensemble of stuctures for the development of a DPM would not adequately represent

Table 3.12: The RMS coordinate deviation values of the six structures selected from clustering of the MD simulation of PfSpdSyn, PDBid 2I7C and the three monomers of PfSpdSyn, PDBid 2PT9, are presented below. The table contains the RMSD values for both the backbone and active site residues for the respective structures.

Protein	RMSD (backbone)	Active Site RMSD
Clus1B	0.000	0.000
Clus2B	1.459	1.857
Clus3B	1.191	1.317
Clus5B	0.717	0.686
Clus1C	1.667	2.240
Clus5C	1.803	2.396
2PT9A	1.555	1.985
2PT9B	1.647	2.167
2PT9C	3.771	3.681

the binding characteristics of the active site and in particular the putrescine binding cavity. Subsequently, it was decided to include the three monomers of PfSpdSyn (PDBid 2PT9) in the negative image construction of the active site of PfSpdSyn.

The six structures from the subensemble of structures were aligned with the three monomers of the PfSpdSyn (PDBid 2PT9) and their RMSD values calculated for both the backbone and active site residues (Table 3.12). It was concluded that these nine structures of PfSpdSyn provide adequate phase space sampling and were used in the negative image construction stage.

3.3.3. Negative Image Construction, Hit Analysis and *In Vitro* Testing

3.3.3.1. MIF Analysis and Pharmacophore Feature Identification

The development of a DPM requires that the chemical space within the active site of the protein is explored to find energetically favourable binding hotspots representing HBA, HBD and HYD characteristics. This is done by performing MIF analysis using various chemical probes featuring such characteristics.

Initially a pilot study was launched to compare PhFs derived from MIF analysis to known binding areas from PfSpdSyn structures co-crystallized with inhibitors. This pilot
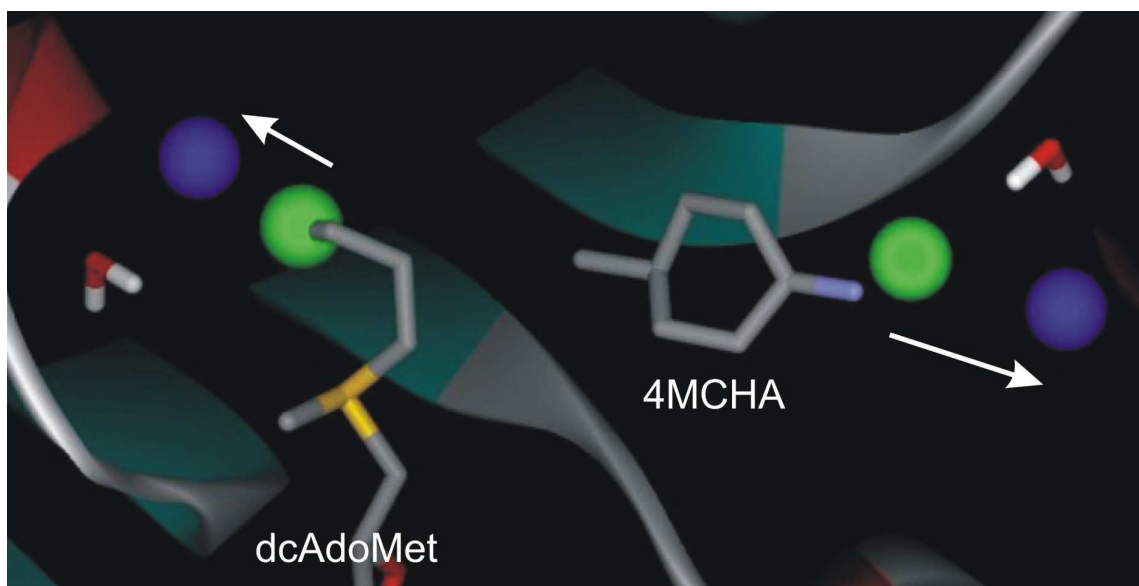
Figure 3.25: An illustration of the geometric shift of two PhFs from two known inhibitors representing the nitrogen of the aminopropyl group of dcAdoMet and the nitrogen of 4MCHA in the absence of water molecules within the active site. In the absence of the two water molecule (shown in sticks) these PhFs (green spheres) shift deeper into their respective binding cavities with their new positions represented by the blue spheres. The direction of the shifts are indicated by white arrows.

study indicated a shift of two known PhFs (binding areas). It showed that the PhFs representing the nitrogen of the aminopropyl group of dcAdoMet and the nitrogen of 4MCHA (non-attacking nitrogen of putrescine), were both shifted more or less 2Å deeper into their respective binding cavities (Figure 3.25). To investigate this a visual inspection of the active site of the PfSpdSyn crystal structures (PDBid 2I7C and 2PT9) revealed that water molecules were co-crystallized within the areas to where the PhFs had shifted. The interactions which these water molecules make with the inhibitors (substrates), explain the shift of both the PhFs.

To correct for this shift MIF analysis using a water probe was performed on the subensemble of protein structures to determine the binding hotspots for these waters. Extra caution was taken during the inclusion of water molecules within the active site, since it is known that the inclusion of water molecules is not straightforward (Garcia-Sosa et al., 2005). Water binding hotspots identified from MIF analysis were used instead of water molecules from the MD simulation, since the MD simulation was performed void of

Table 3.13: The coordinates and energies of the water molecules added to the subensemble of PfSpdSyn structures, which were used during GRID analysis.

Cluster	Coordinates			Additional Information		
	x	y	z	rgyr	ave beta	Comments
Clus1B	2.430	0.709	-10.530	0.43	-12.50	Aminopropyl Nitrogen
Clus1B	1.272	7.823	6.610	1.01	-11.71	Non-attacking Nitrogen
Clus2B	1.534	1.406	-8.436	1.01	-12.36	Aminopropyl Nitrogen
Clus2B	Not Identified					Non-attacking Nitrogen
Clus3B	2.125	0.642	-10.071	0.78	-12.10	Aminopropyl Nitrogen
Clus3B	Not Identified					Non-attacking Nitrogen
Clus5B	1.843	0.734	-10.400	0.91	-12.10	Aminopropyl Nitrogen
Clus5B	Not Identified					Non-attacking Nitrogen
Clus1C	1.716	0.889	-10.555	1.34	-11.95	Aminopropyl Nitrogen
Clus1C	3.233	6.696	7.116	0.55	-13.17	Non-attacking Nitrogen
Clus5C	1.023	2.021	-8.902	0.90	-11.34	Aminopropyl Nitrogen
Clus5C	Not Identified					Non-attacking Nitrogen

inhibitors and it is thought that MIF analysis would give a better representation of the water molecules and their respective binding hotspots. Binding hotspots for the water molecules of interest were identified for all the structures in the subensemble of PfSpdSyn and added to the respective structures. The 3D coordinates and interaction energies of the respective water molecules added are shown in Table 3.13.

The representative structures of clus2B, Clus3B, Clus5B and Clus5C were shown not to have binding hotspots for water molecules within the non-attacking nitrogen of putrescine binding area and were thus not included. Both Clus1B and Clus1C structures showed water binding hotspots within the non-attacking nitrogen binding area and had strong binding energies ranging between -11.7 and -13.1 kcal/mol and were therefore included. These water molecules facilitated binding of ligands and can be seen in the crystal structures of PfSpdSyn co-crystallized with AdoDATO, 4MCHA and spermidine (PDBids; 2I7C, 2PT9 and 2PWP, respectively). When performing MIF analysis these water molecules effectively act as a stopper molecule within the active site, facilitating the analysis by providing HBD and HBA characteristics within the binding areas of interest. Water molecules can be displaced upon binding of a ligand within the active site and the combination of structures with and without water molecules can be considered to give a
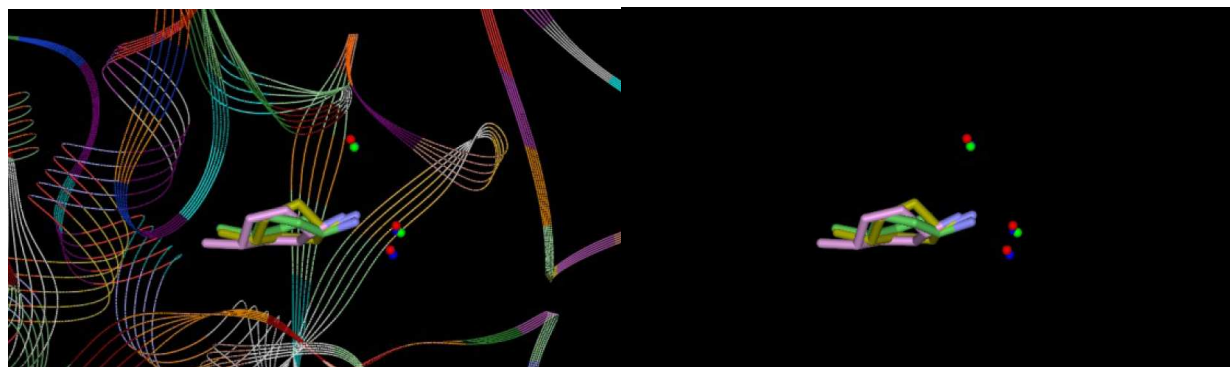
Figure 3.26: In the crystals structure of PfSpdSyn (PDBid 2PT9) water molecules interacting with 4MCHA are not present in all the monomers crystallized. Water molecules are represented as spheres in blue (chain A), green (chain B) and red (chain C).

good representation of the active site and its binding capability in its natural environment.

The two crystal waters corresponding to the water molecules identified in the non-attacking nitrogen binding region as discussed above were included during MIF analysis for PfSpdSyn (PDBid 2PT9). These two water molecules have been proposed by Dufe *et al.* (2007) to play a role in the stabilization of 4MCHA, when bound to PfSpdSyn (PDBid 2PT9) and to form hydrogen bonds with Glu 231 and Glu 46, respectively (Dufe *et al.*, 2007). However, these water molecules were not found to be present in all the PfSpdSyn (PDBid 2PT9) monomers crystallized (Figure 3.26). It should be noted from Figure 3.26 that a water molecule from chain B (green) of PfSpdSyn (PDBid 2PT9) is missing. The missing water molecule supports the previous findings where no water binding hotspots were found in some of the structures captured during MD simulations. It can, however, not be excluded that the missing water molecule was due to experimental difficulties in crystallization.

The incorporation of additional target flexibility while performing MIF analysis is possible by changing the Move parameter of the GRID program. A comparison between MIF analysis results of the selected subensemble of PfSpdSyn structures was made using both static and flexible target structures (Figure 3.27). The comparison showed that known binding areas were not identified when using a flexible target structure. For

example, Figure 3.27 illustrates the difference in grid points between a static (blue grid points) and flexible (red grid points) target structure, using an N3+ probe (NH$_3^+$). From Figure 3.27, it is clear that when using a flexible target structure no grid points are shown representing the binding area of the attacking nitrogen of putrescine (yellow wire-frame). It is known that the incorporation of flexibility of the target structure during MIF analysis is notoriously difficult. Quoting the GRID documentation (`http://www.moldiscovery.com/docs/grid/`) "It must be emphasized that this three-stage model is only a model (The real-life process is presumably very much more complicated). However, stages 1 and 2 are relatively easy to compute on the basis of this model, while stage 3 is much more difficult. We hope this Version of the program Grid does a reasonable job with the first two stages, and that it does not do a very bad job with stage three". Considering this, it was decided not to include the additional flexibility of the target structure when performing MIF analysis.
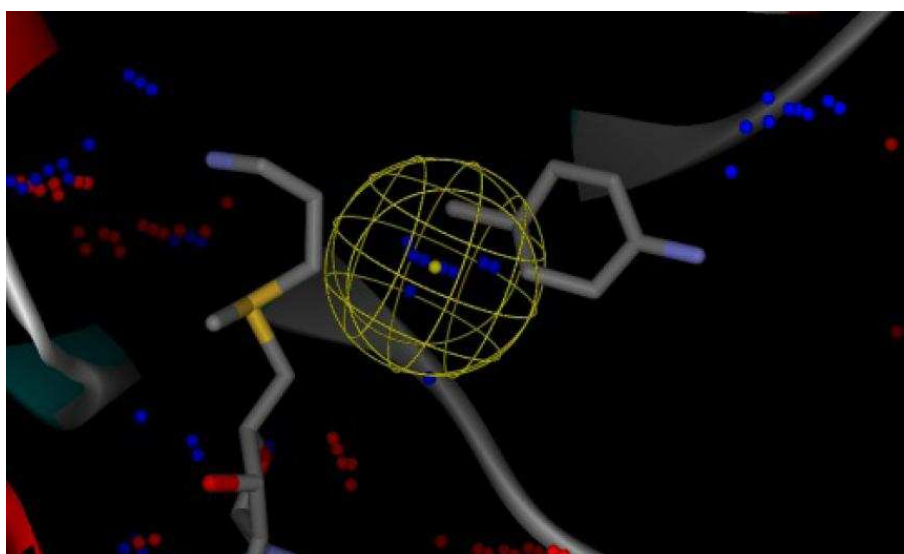

Figure 3.27: A comparison between the results obtained from a MIF analysis using either a static (Move = 0; Blue grid points) or flexible (Move = 1; Red grid points) structure. The yellow wire-frame sphere represents the binding area of the attacking nitrogen from putrescine. The residues presented, Asp 197, Ser 198 and Tyr 263 are from Cluster 1 (chain B; Green cartoon representation). From this it is clear that MIF analysis with a flexible target does not find the binding area of the attacking nitrogen from putrescine.

Prior to the selection of probes for MIF analysis, known inhibitors and substrates

were evaluated to identify chemical moieties representing known HBA, HBD and HYD characteristics. The most common chemical moieties were found to be the NH_3^+, OH, CH_3 and NH_2 entities and were considered in the selection of probes to explore the PfSpdSyn active site. The probes selected to represent HBD characteristics were the N1+, N2 and N3+ probes. The N1+ probe represents a sp^3 amine (NH) cation, the N2 probe represents a neutral flat nitrogen with two hydrogens (NH_2), used to mimic amino groups whereas the N3+ probe represents a sp^3 amine (NH_3 cation), used to mimic the amine cation entities of polyamines. The probes selected to represent the HBA characteristics were the O, O1 and N:= probes. The O probe represents a sp^2 carbonyl oxygen, the O1 probe represents an alkyl hydroxy (OH) group and the N:= probe a sp2 nitrogen with lone pair. The HYD characteristics were represented by the C3 and DRY probes. The C3 probe represents a methyl CH_3 chemical moiety, whereas the DRY probe is a general hydrophobic probe.

Grid points obtained after MIF analysis were extracted and displayed within the active site of PfSpdSyn. Clusters of grid points were identified by visual inspection and their center of mass calculated and weighted by binding energy, to give a better representation of the true binding hotspots. Also the radius of gyration was calculated to estimate the size of the PhF. This was done for all the probes.

3.3.3.2. Exclusion volumes

Exclusion volumes were selected to facilitate the database searching. Atoms within 7Å of AdoDATO in the PfSpdSyn crystal structure, PDBid 2I7C, were selected. The atom coordinates were then extracted from all the structures present in the PfSpdSyn subensemble and their center of mass calculated to incorporate the target flexibility in the EVs. These "dynamic" EVs give a better representation of the boundaries of the active site, since it gives an average position of the atoms over time (MD simulation) and incorporates information from all the structures of the PfSpdSyn subensemble. The EVs for PfSpdSyn were overlaid with the co-crystallized ligands AdoDATO, dcAdoMet and 4MCHA within the active site to identify and remove clashes. For example, clashes were found between AdoDATO (also 4MCHA) and Gln 229 (Gln 229 was discussed in Section

3.3.2.2). The EVs represented by atoms OE1, NE2 and CD of Gln 229, clashed and were removed, since they filled spaces that are occupied during ligand binding (e.g. AdoDATO and 4MCHA). The entire active site was investigated and all clashes removed. The EVs radii were changed to 50% of its original value (1.2Å) to increase the search-able chemical space within the active site. The EVs selected were added to the DPMs to facilitate database searching.

3.3.3.3. Pharmacophore Model Selection

The active site of PfSpdSyn is divided into two binding cavities, one for putrescine and one for dcAdoMet (Haider *et al.*, 2005; Ikeguchi *et al.*, 2006). Figure 3.28(a) shows a 2D representation of the PfSpdSyn active site with the natural occurring substrates in their respective binding cavities. To facilitate the search for new inhibitors, the active site was further sub-divided into four alternative binding regions, DPM1 through to DPM4 (Figure 3.28). The DPM1 binding region was used to explore the putrescine binding cavity (Figure 3.28(b) (green)). The DPM2 binding region was selected to explore compounds which protrude from the putrescine binding cavity into the dcAdoMet cavity by bridging the catalytic center (Figure 3.28(c) yellow). The DPM3 binding region included the catalytic center and was used to explore the dcAdoMet binding cavity (Figure 3.28(d) dark blue). The DPM4 binding region was used to explore the entire active site of PfSpdSyn (Figure 3.28(e) red). For the sake of clarity, DPM1 to DPM4 binding regions will be referred to as binding cavities in the following sections.

Due to the knowledge-based nature of this project compounds were identified not only by the DPM methodology but also by rationally deriving compounds and incorporating knowledge gained at various stages of this study. Illustrated in Figure 3.29 are the two methodologies followed to derive compounds for *in vitro* testing.

3.3.3.4. DPM1 Binding Cavity

The DPM1 binding cavity was selected specifically to explore and identify novel PhFs in the putrescine binding cavity, which can be used together with PhFs derived from protein-ligand complexes to search for compounds that binds more strongly within this
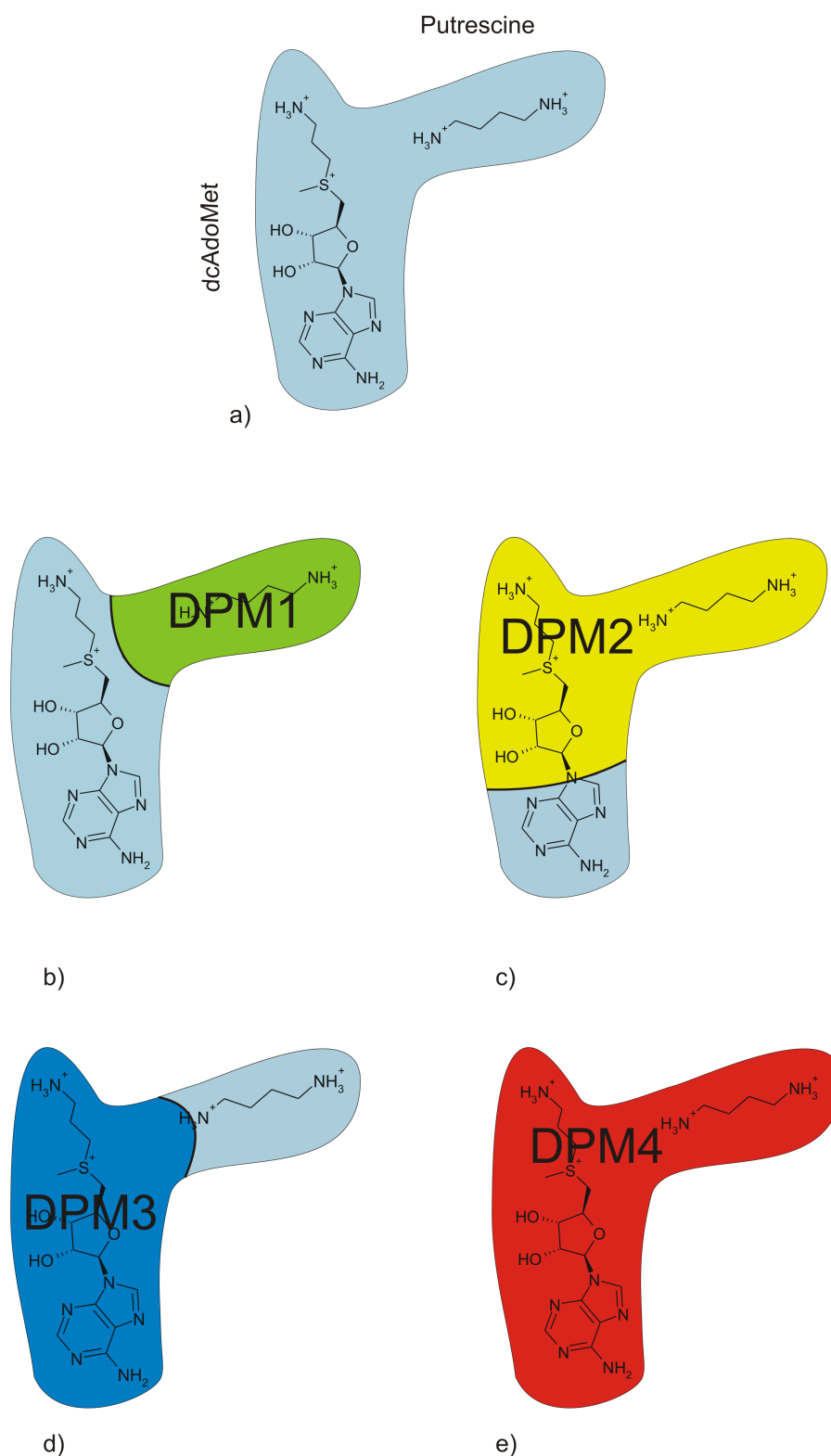
Figure 3.28: A 2D representation of the active site of PfSpdSyn illustrating different regions used to explore and construct pharmacophore models. a) The PfSpdSyn active site containing the natural occurring substrates, dcAdoMet and putrescine within their respective binding cavities. b) Illustrated in green is DPM1 which was used to explore the putrescine binding cavity. c) DPM2 is illustrated in yellow and is an extension of DPM1 extending into the dcAdoMet cavity. d) DPM3 in dark blue illustrates the part of the active site that was used to explore the dcAdoMet cavity, including the catalytic center. e) DPM4 (illustrated in red) was used to explore both the dcAdoMet and putrescine binding cavities simultaneously.
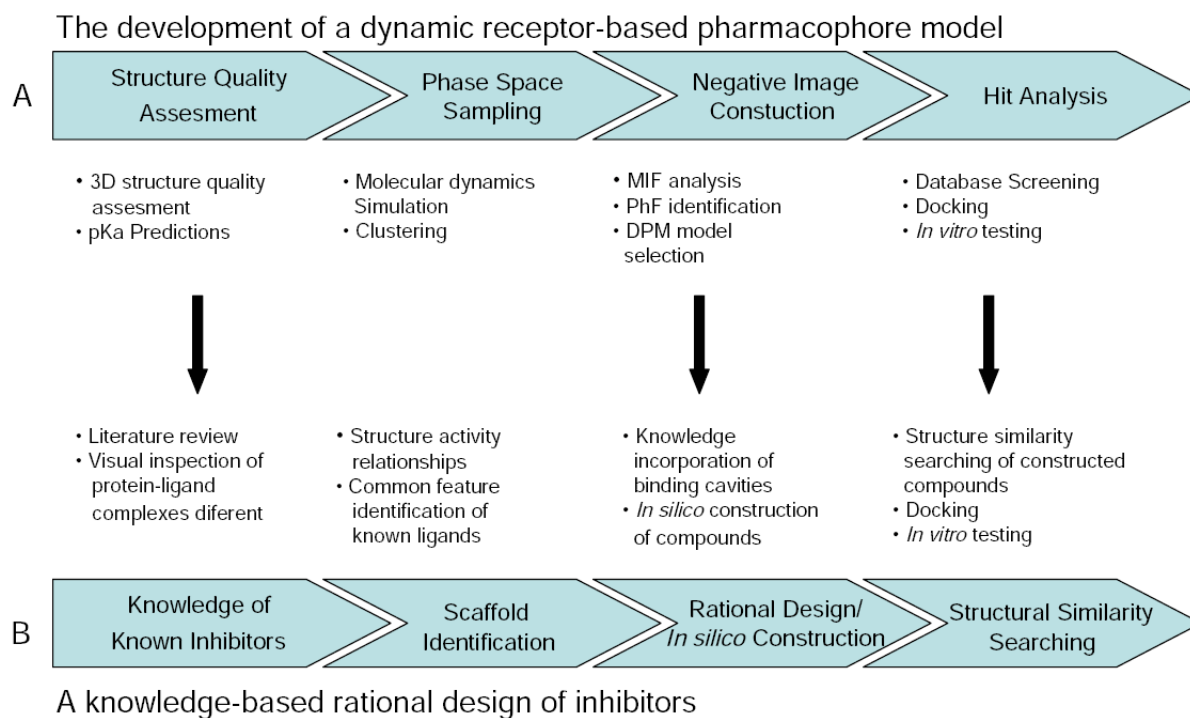
Figure 3.29: Two methodologies followed to derive inhibitors for *in vitro* screening against PfSpdSyn. *(Top)* The four major stages in the development of a dynamic receptor-based pharmacophore model for PfSpdSyn, which are subdivided into the major steps followed to identify inhibitors. *(Bottom)* The four major stages involved in rationally deriving inhibitors for PfSpdSyn using a knowledge-based design approach. These stage were subdivided into the major steps followed at each stage.

cavity. Before continuing with the studies of the DPM1 binding cavity, two PhFs were tested to verify that they can be used to identify 4MCHA, an inhibitor which binds within the putrescine binding cavity. The two PhFs used to identify 4MCHA consisted of one hydrophobic and one positive ionizable (PI) PhF and were derived from the PfSpdSyn (PDBid 2PT9) crystal structure (Figure 3.30). The hydrophobic PhF was selected to represent the methyl group of 4MCHA and the PI group, the protonated nitrogen. 4MCHA has a best fit value of 0.998/2 with this pharmacophore model. It should be noted that fit values are a function of the number of PhFs within the model, therefore in this study fit values are given with the number of PhFs within the pharmacophore model (e.g. 0.998/2; 0.998 the fit value and 2 the number of features within the pharmacophore model). A higher fit value represents a better superimposition of the compound's functional groups
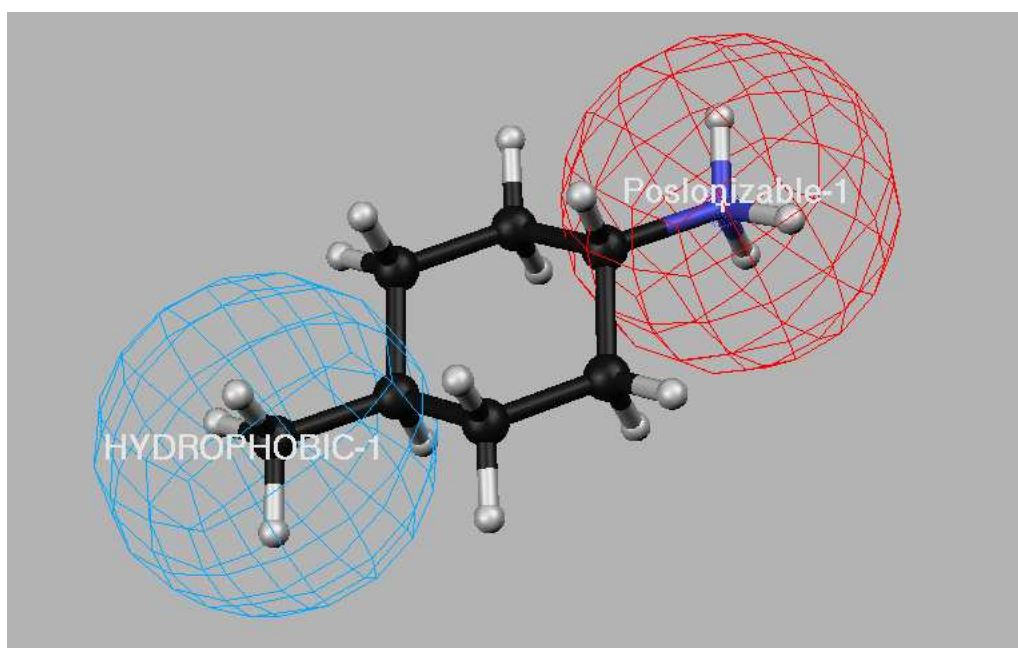
Figure 3.30: 4MCHA fitted to a pharmacophore model represented by a hydrophobic (blue) and positive ionizable (red) PhF (Fit Value 0.998/2).

to a pharmacophore model.

The PhFs derived from MIF analysis were used to construct a shortlist of DPMs for the DPM1 binding cavity. A total of eight DPMs were constructed (DPM1a-DPM1h) and consisted either of three or four PhFs each. The PhFs were selected from both the subensemble of structures derived from the molecular dynamic simulation of PfSpdSyn and those identified from the PfSpdSyn crystal structure, PDBid 2PT9. PhF selection for the DPMs of the DPM1 binding cavity were focused on the PhFs derived from the crystal structures, since it is believed that the PhFs give a better representation of the putrescine binding cavity as discussed in section 3.3.2.2.

The eight DPMs derived from the DPM1 binding cavity (DPM1a-h; Figure 3.31) were all screened against the drug-like subset of the ZINC database. These database screens resulted in the identification of between 100 and 22 000 hits, which needed further evaluation by docking studies. The docking of a compound to a receptor is a more computationally expensive procedure than screening and therefore only a limited number of compounds
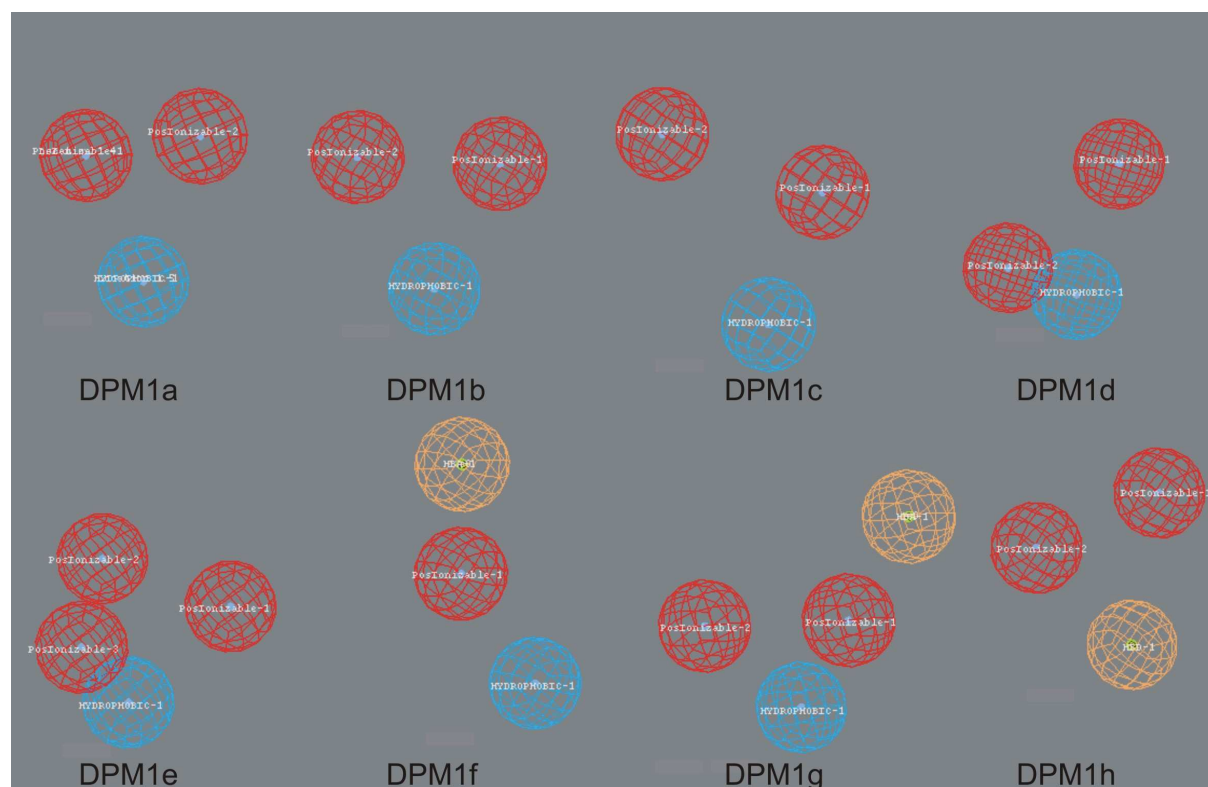
Figure 3.31: The eight DPM1's (DPM1a-h) selected to screen the drug-like subset of the ZINC database. Spheres in red represent positive ionizable groups, those in blue represent the hydrophobic features and those in orange either hydrogen bond donors or acceptors.
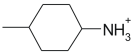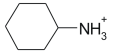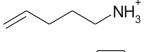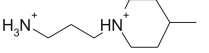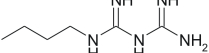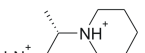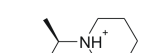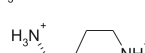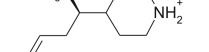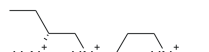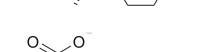
can be docked. The standard procedure for filtering down the number of compounds is to calculate the best fit values of the compounds to the DPMs and to select the compounds accordingly. This strategy was adjusted for the current project since the calculation of best fit values of 22 000 compounds is comput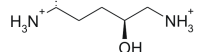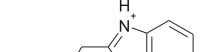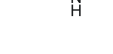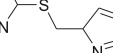ationally too expensive. Therefore the compounds resulting from the database screens were sorted according to their molecular weight (MW) in an ascending order, thereby excluding compounds exceeding a certain MW (<400 Da) that would not be able to occupy this site. Smaller compounds could result in cooperative binding with dcAdoMet and/or MTA which enhances the potency of the compounds as was shown with 4MCHA and APE (Dufe *et al.*, 2007). In addition a visual inspection of all compounds thought to possibly bind in the putrescine binding cavity was also made. Before selecting the compounds a best fit value was calculated and used as a further selection criterion. From the visual inspection of the DPM1 binding cavity searches, eleven compounds were selected and included the known inhibitors of the

DPM1 binding cavity (Table 3.14).

The eleven compounds selected were subjected to docking using AutoDock 4.0 to further filter down and rank the compounds based on their binding poses and docking energies within the active site. The compounds were docked to two crystal structures of PfSpdSyn, PDBid 2I7C and PDBid 2PT9 to address a common problem found in docking studies, the cross-docking problem (Krovat *et al.*, 2005). For example 4MCHA and dcAdoMet could be docked correctly back into PDBid 2PT9 (RMSD 1.29Å and 0.69Å, respectively). However, the inhibitor AdoDATO co-crystallized with PfSpdSyn, could not be docked to the PfSpdSyn crystal structure, PDBid 2PT9. On the contrary, it was possible to dock 4MCHA and dcAdoMet to PfSpdSyn, PDBid 2I7C. Therefore, all compounds selected from the database screens were docked to both the above-mentioned PfSpdSyn structures in an attempt to accommodate the cross-docking problem. The compounds were not docked to the subensemble of protein structures sampled during the MD simulation since the conformational changes that Gln 229 undergo upon binding of inhibitors, change the characteristics of the putrescine binding cavity significantly (section 3.3.2.2). Water molecules were excluded in the docking of any of the compounds, since their orientation within the active site has significant consequences on the docking and scoring procedures, and thus the ranking and selection of compounds (Krovat *et al.*, 2005). The compounds selected from the DPM screens describing the DPM1 binding cavity were docked in the presence of dcAdoMet, thus preventing their crossing into the dcAdoMet binding cavity. The eleven docked compounds were visually analyzed by examining their binding poses and docking energies (Table 3.14).

Compounds were selected to avoid overlapping of chemical features e.g. compounds ZINC02582060, ZINC04226830 and ZINC04226831 are very similar, therefore only one compound of the three needed to be selected to test for specific features or a certain substructure (Table 3.14). All three of the compounds share a common core structure and docking of these compounds shows good binding energies as well as good binding poses. The docked ZINC02582060 showed hydrogen bond formation with Asp 199 and Glu

Table 3.14: Compounds used in the exploration of the DPM1 binding region. These compounds were identified from either DPM screens against the drug-like subset of the ZINC database or rationally derived. The DPMs are representative of the chemical characteristics found within the DPM1 binding region.

Compound id	Structure	Best Fit Value/ number of PhFs	DPM model	Docking Energies PDBid 2I7C	PDBid 2PT9
trans-4MCHA		0.99/2	Test	-7.2	-7.2
CHA		N/A	—	-6.7	-6.7
APE		N/A	—	-5.5	-5.6
ZINC02582060		1.56/3	DPM1b	-9.2	-9.4
ZINC04097425		1.71/3	DPM1b	-7.74	-7.7
ZINC04226830		1.56/3	DPM1b	-9.3	-9.9
ZINC04226831		1.18/3	DPM1b	-9.4	-9.7
ZINC04716451†		1.86/3	DPM1b	-9.8	-9.3
ZINC04716453		1.86/3	DPM1b	-8.7	-9.1
ZINC04226723†		1.68/3	DPM1d	-10.1	-10.3
ZINC01532890		0.57/3	DPM1h	-8.0	-8.0
ZINC02025639		0.09/3	DPM1h	-9.6	-8.7
ZINC03790677		0.07/3	DPM1h	-5.2	-5.1
ZINC04512857		0.15/3	DPM1h	-5.5	-5.2
2AMCHA		N/A	—	-8.5	-9.2
4AMP		N/A	—	-7.6	-7.7

†No good binding pose was found

N/A Not Applicable since these compounds were not derived using pharmacophore models

231 (Figure 3.32). The same hydrogen bond pattern formation was shown for the docked ZINC04226830 and ZINC04226831 compounds (Table 3.14). The main difference between these three compounds is the side chains attached to the piperidine ring. ZINC02582060 was selected above ZINC04226830 and ZINC04226831 since it was expected that the propan-1-amine chains of the latter two compounds would have unfavourable interactions with PfSpdSyn when compared to the aminopropyl chain of ZINC02582060.

A further selection criteria to test ZINC02582060 *in vitro* was its similarity in structure and docking pose to 4MCHA (Figure 3.32). However, the docking energy of ZINC02582060 to PfSpdSyn was lower than that of 4MCHA (Table 3.14). The aminopropyl chain of ZINC02582060 forms hydrogen bonds with residues Asp 199 and Glu 231 and another hydrogen bond was formed between residue Asp 199 and the nitrogen of the piperidine ring (Figure 3.32). The similar docking pose of ZINC02582060 to 4MCHA is expected to provide information on whether the putrescine binding cavity is flexible enough to accommodate the additional aminopropyl chain protruding from the piperidine ring (Figure 3.32). ZINC02582060 has an IC_{50} value of 0.5 μM in the *trans* configuration when tested as a spermine synthase inhibitor of *Rut rattus* (Shirahata *et al.*, 1991). Taken together, this provides confidence in the selection of compound ZINC02582060, since PfSpdSyn can accommodate spermine as a substrate within the putrescine binding cavity (DPM1; Haider *et al.* 2005).

Jacobsson *et al.* (2008) reported during the course of this study that ZINC02025639 (also identified in this study) to be an inhibitor of PfSpdSyn and able to bind in the putrescine binding cavity (no K_i and no IC_{50} values were given; Table 3.14). The docking pose of ZINC02025639 using AutoDock was found to be the same as proposed by Jacobsson *et al.* (2008). It was also observed with NMR experiments that cooperative binding exists between MTA and compound ZINC02025639 leading to the proposal that the binding affinity of ZINC02025639, which has a high IC_{50} of 159μM to PfSpdSyn, is similar to that of MTA. This compound was therefore not selected for further studies. The identification of ZINC02025639 in this study demonstrated that the selected DPMs
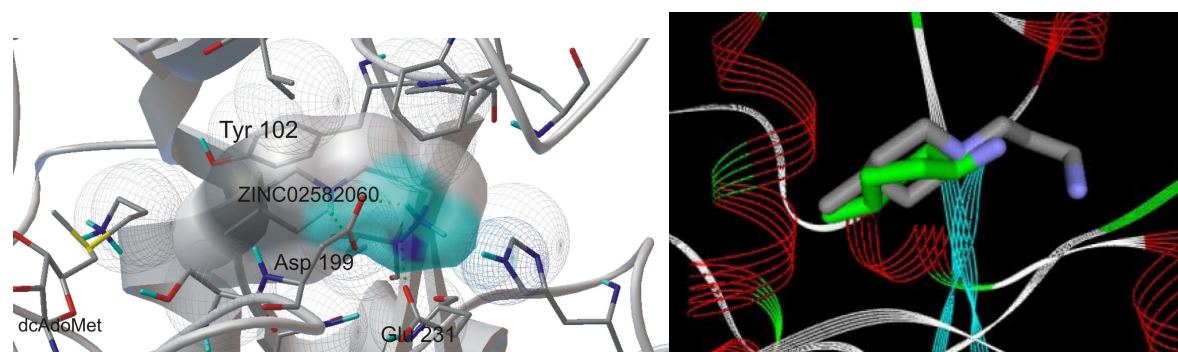
Figure 3.32: An illustration of the docking pose compound ZINC02582060 assumes when bound to PfSpdSyn (PDBid 2PT9). *(Left)* Compound ZINC02582060 docked to PfSpdSyn. Illustrated with green dots are the hydrogen bonds formed with residues Asp 199 and Glu 231 of PfSpdSyn. *(Right)* The docked compound ZINC02582060 overlaid with 4MCHA (green) co-crystallized with PfSpdSyn (PDBid 2PT9).

were able to identify the compound within the putrescine cavity albeit at a low fit value.

A third compound, *trans*-4-Amino-1-methylpiperidine (4AMP) was selected to be tested *in vitro* for further exploration of specific characteristics of the putrescine binding cavity (DPM1). During database searching a large percentage of high-scoring compounds (either based on fit values or docking energies) were identified containing a piperidine ring. It is well known that the putrescine binding cavity (DPM1) can accommodate cyclohexane rings but, no data are available on its ability to accommodate piperidine rings. Such information would be invaluable in the selection process of compounds to be tested as inhibitors of PfSpdSyn (e.g ZINC02582060 and ZINC01683322 (DPM2)). 4AMP was identified via a structural similarity search of 4MCHA using SciFinder and was selected to determine the effect which the piperidine's ring nitrogen would have upon binding to PfSpdSyn. 4AMP is a weak substrate (<1% MTA production compared to putrescine) of aminopropyltransferases (SpdSyn and SpmSyn; Shirahata *et al.* 1991). It is not known whether it also inhibits these enzymes. The selection of 4AMP is therefore based on the assumption that it would bind to PfSpdSyn in a similar binding pose as 4MCHA and that the effect of the replacement of cyclohexyl ring with a piperidine ring would be reflected in the inhibition of PfSpdSyn. Docking of 4AMP to PfSpdSyn shows that the piperidine ring nitrogen is accommodated within the putrescine binding cavity

(DPM1). However, the ring is adjusted within this cavity with respect to the binding orientation of 4MCHA (Figure 3.33). Since the *trans* configuration of 4AMP was docked to PfSpdSyn, 4AMP was overlayed to the docked *trans*-4MCHA configuration (PDBid 2PT9) and not the crystallized *cis*-4MCHA in order to remove docking bias with regards to orientation (Figure 3.33).
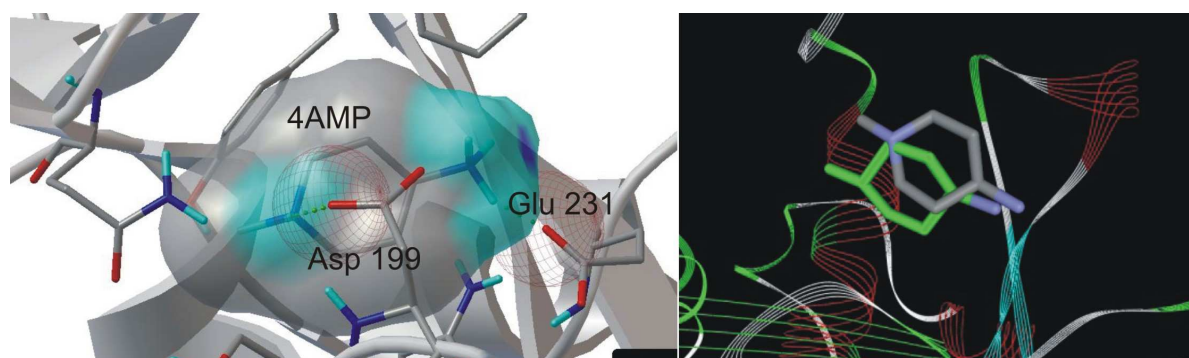

Figure 3.33: Results from the docking of 4AMP to PfSpdSyn (PDBid 2PT9). *(Left)* The docking pose and hydrogen bonds 4AMP makes when docked to PfSpdSyn. *(Right)* An overlay of *trans*-4MCHA and *trans*-4AMP within the putrescine binding cavity after docking.

Both ZINC02582060 and 4AMP were subsequently tested *in vitro* against the activity of PSpdSyn. The *in vitro* testing was performed by S. Reeksting during a research visit to the laboratory of Prof R.D. Walter at the Bernard Nocht Institute for Tropical Medicine, Hamburg, Germany. The method followed to perform the *in vitro* testing was described by S Reeksting as part of his *MSc* study (Haider *et al.*, 2005; Reeksting, unpublished). ZINC02582060 and 4AMP at a 100 μM concentration showed 18% and 5% reduction of the PfSpdSyn activity, respectively (Figure 3.34). By comparison, inhibition by 4MCHA of PfSpdSyn as control resulted in a 94% activity loss (Figure 3.34).

The *in vitro* data obtained for 4AMP suggests that its piperidine ring could not be accommodated within the putrescine binding cavity. However, the piperidine ring of ZINC02582060 seems to be accommodated within the PfSpdSyn putrescine binding cavity as suggested by the *in vitro* data. From docking studies it was proposed that 4AMP binds with its piperidine ring nitrogen oriented towards the catalytic center (Figure 3.33) whereas the ring nitrogen of ZINC02582060 is oriented towards the non-attacking side of
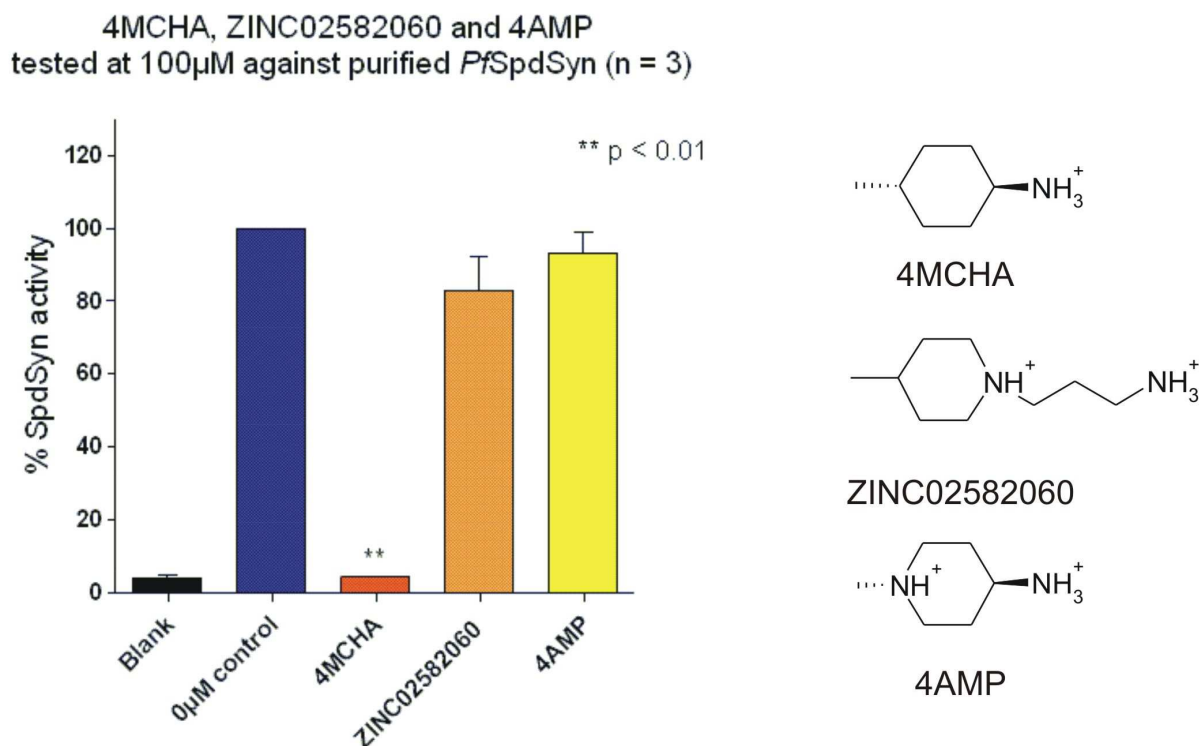
Figure 3.34: Inhibitory effect of 100 μM of 4MCHA, ZINC02582060 and 4AMP on PfSpdSyn. *(Left)* A bar graph depicting the reduction in PfSpdSyn activity of 4MCHA, ZINC02582060 and 4AMP. *(Right)* The chemical structures of the compounds tested *in vitro*.

the putrescine binding cavity (Figure 3.32). Although inhibition by ZINC02582060 is not very potent it is proposed that further studies need to be performed such as e.g. reducing the carbon length of the aminopropyl chain to determine whether or not the inhibition can be improved. The data of obtained from *in vitro* testing of 4AMP and ZINC02582060 suggest that a piperidine ring can be accommodated within the putrescine binding cavity depending on its orientation with regards to its ring nitrogen.

The development of a DPM for PfSpdSyn provides a wealth of information regarding its active site, which can be used in the modification of existing inhibitors. MIF analyses provide binding hotspots within the active site represented by PhFs. These PhFs were overlaid within the putrescine binding cavity to identify possible modifications that could be made to 4MCHA to increase its binding affinity. In this section a knowledge-based approach was used to optimize existing inhibitors using information gained from MIF

analysis. In Figure 3.35 two PhFs representing the HYD (methyl group) and PI (nitrogen) features of 4MCHA are presented. A third feature identified using MIF analysis was found to have the lowest energy value (-26.1 kcal/mol) of the PhFs identified within the putrescine binding cavity. The chemical space where this feature was identified is usually occupied by a water molecule that has been shown to be replaceable upon binding of spermine (PDBid 3B7P). It is therefore thought that this water can be replaced by a PI group and was further investigated.
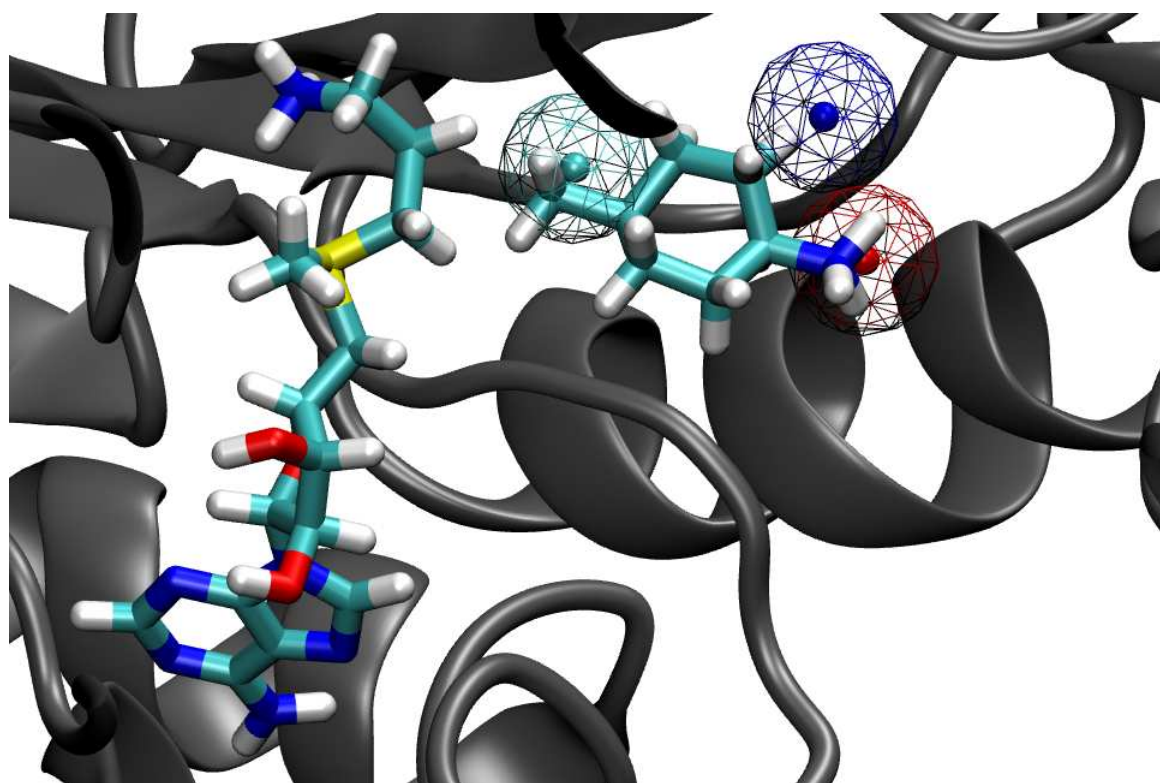

Figure 3.35: Identification of a new binding hotspot within the putrescine binding cavity of PfSpdSyn (PDBid 2PT9). In the figure are displayed dcAdoMet and 4MCHA. The red mesh sphere represents the PI PhF identified for 4MCHA and the cyan mesh sphere the hydrophobic PhF. In blue is represented a new PI PhF (binding hotspot) in close proximity to 4MCHA.

A substructure search was performed against the drug-like subset of the ZINC database using 4MCHA as a reference structure (scaffold) and allowed substitutions on the ortho and meta positions. A similar search was performed using CHA allowing for any substitution on the ortho, meta and para positions. The results did not deliver any compounds that were not identified during the DPM searches. Subsequently, structure similarity

searching was performed using both PubChem and SciFinder but focusing on the commercial availability of compounds. *trans*-2-Aminomethyl-cyclohexylamine (2AMCHA; Table 3.14) was identified as a possible inhibitor by performing a structure similarity search using SciFinder. 2AMCHA was docked to PfSpdSyn and found to have a similar orientation to that of 4MCHA (Figure 3.36).
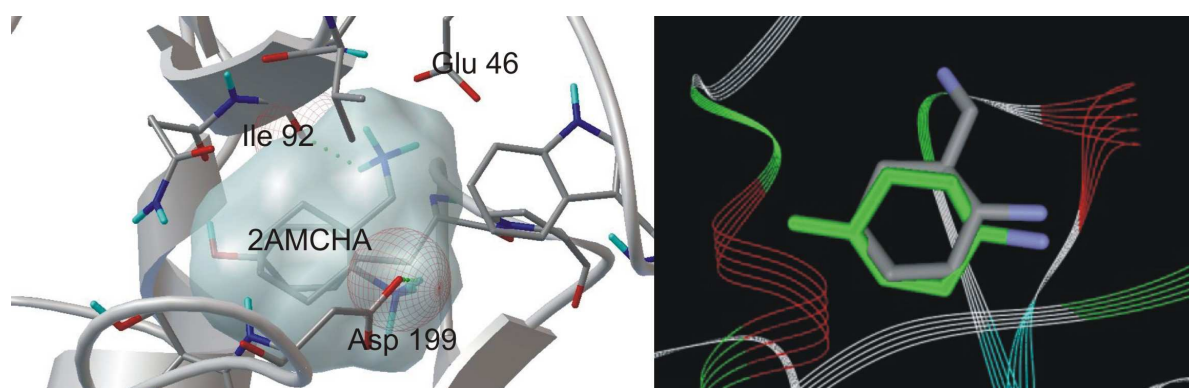

Figure 3.36: An illustration of the docking pose 2AMCHA assumes when docked to PfSpdSyn (PDBid 2PT9). *(Left)* The docking pose and hydrogen bond interactions of 2AMCHA when docked to PfSpdSyn. *(Right)* An overlay of 4MCHA and 2AMCHA docked to PfSpdSyn (PDBid 2PT9).

The docking study revealed that the aminomethyl group of 2AMCHA protrudes into the PI binding hotspot identified (Figure 3.35) which may explain the increase in the binding affinity above that of 4MCHA and CHA. Since 2AMCHA show some structural similarity to putrescine, the possibility to act as a substrate can not be excluded. Therefore, a methyl group was added at the para position *in silico* to evaluate if it would have a significant effect on the orientation and binding of the proposed 2AMCHA after docking. The docking orientation was found to be very similar to that of 2AMCHA (data not shown). The possibility, however, can not be excluded that 2AMCHA may act as a substrate when bound to PfSpdSyn. In summary, 2AMCHA was selected to explore a newly identified PI binding hotspot within the putrescine binding cavity. Unfortunately compound 2AMCHA could not be obtained at the time of *in vitro* testing but will be included in future studies.

3.3.3.5. DPM2 Binding Cavity

The DPM2 cavity was selected to explore the chemical space extending from the putrescine binding cavity into the dcAdoMet binding cavity by bridging the catalytic center (Figure 3.28b). At the moment, no inhibitors are known to bind in this cavity. PhFs within this binding cavity were specifically selected and fourteen DPMs were constructed (DPM2a-DPM2n; Figure 3.37). As with the DPM1 cavity, the majority of the PhFs representing the putrescine binding cavity were selected from the monomers of PfSpdSyn, PDBid 2PT9. The DPMs were subsequently screened against the drug-like
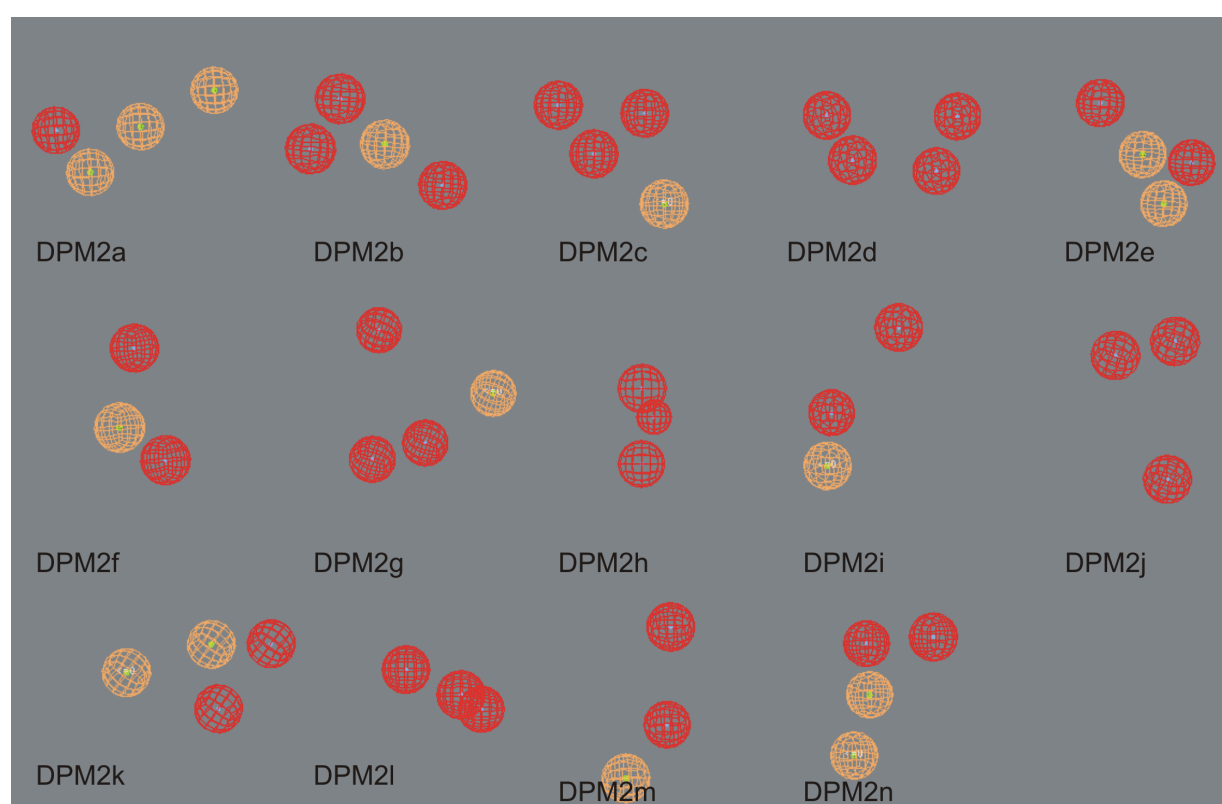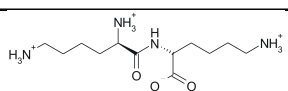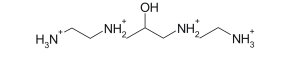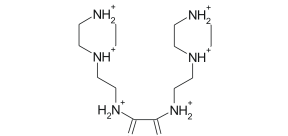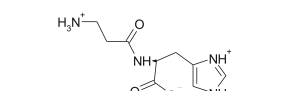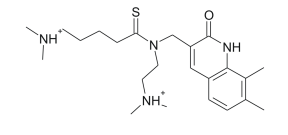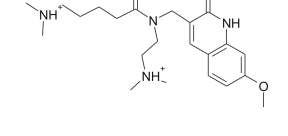

Figure 3.37: Fourteen DPMs (DPM2a-n) representing the most important PhFs identified within the DPM2 binding cavity. Spheres in red represent the positive ionizable groups and those in orange represent either hydrogen bond donors or acceptors.

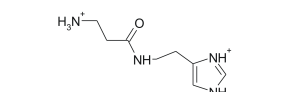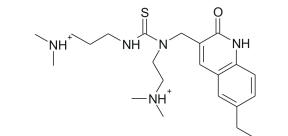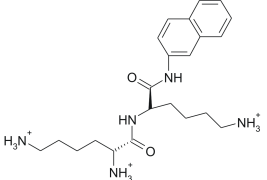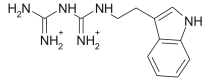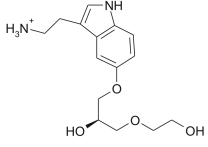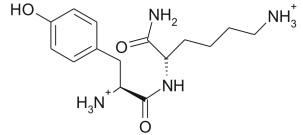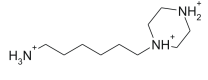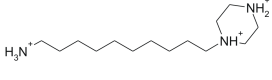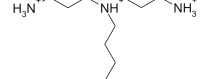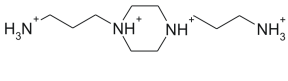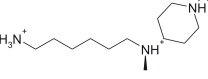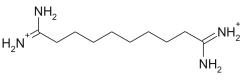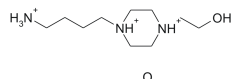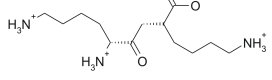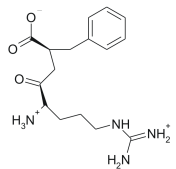subset of the ZINC database. These database screens resulted in 1 to 1800 hits for which the best fit values were calculated. The best fit values were used in combination with visual inspection to select compounds to be used in docking studies. Twenty seven compounds were selected and docked against the same PfSpdSyn structures used in the docking studies of the compounds selected for the DPM1 binding cavity (Table 3.15). No
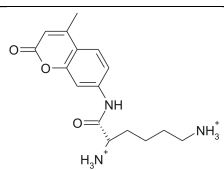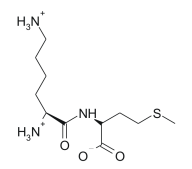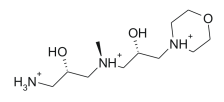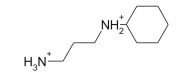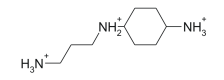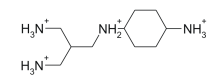
water molecules or ligands were present within the binding cavities during the docking studies.

Before selection of compounds to be tested *in vitro*, all the docked compounds were visually inspected to evaluate their docking poses and related docking energies (Table 3.15).

Table 3.15: Compounds used in the exploration of the DPM2 binding region. These compounds were identified from either DPM screens against the drug-like subset of the ZINC database or rationally derived. The DPMs are representative of the chemical characteristics found within the DPM2 binding region.

Compound id	Structure	Best Fit Value/ Number of PhFs	DPM model	Docking Energies	
				PDBid 2I7C	PDBid 2PT9
ZINC02043137		0.829/4	DPM2a	-12.1	-8.7
ZINC04530101		0.204/4	DPM2a	-12.9	-11.2
ZINC04578206		0.499/3	DPM2a	NBP	NBP
ZINC02040855		0.231/3	DPM2b	-7.58	-6.9
ZINC03871886		0.433/4	DPM2b	NBP	-8.1
ZINC04133443		0.350/4	DPM2b	NBP	NBP
ZINC04133498		0.052/4	DPM2b	NBP	NBP
ZINC04161998		0.172/3	DPM2b	-9.2	NBP

Compound id	Structure	Best Fit Value/ Number of PhFs	DPM model	Docking Energies	
				PDBid 2I7C	PDBid 2PT9
ZINC04899780		0.692/4	DPM2d	-11.7	NBP
ZINC00477084		2.929/4	DPM2c	-11.1	-10.20
ZINC01823645		3.637/4	DPM2c	NBP	NBP
ZINC02561129		3.309/4	DPM2n	-13.1	-11.8
ZINC04204184		0.056/4	DPM2d	-12.7	-10.7
ZINC04204188		0.795/4	DPM2d	-12.8	-11.8
ZINC04977310		0.897/4	DPM2d	-12.5	-10.8
ZINC01683322		0.357/4	DPM2e	-14.0	-12.3
ZINC04226804		0.566/4	DPM2e	-12.2	-11.3
ZINC01762415		2.050/3	DPM2f	-11.0	-10.4
ZINC04226762		1.669/3	DPM2f	-12.1	-10.3
ZINC04556737		0.870/3	DPM2h	-11.1	-10.2
ZINC01607827		0.417/4	DPM2n	-10.7	-9.7

Compound id	Structure	Best Fit Value/ Number of PhFs	DPM model	Docking Energies	
				PDBid 2I7C	PDBid 2PT9
ZINC02517186		0.855/4	DPM2n	-10.5	NBP
ZINC02522672		0.664/4	DPM2n	-8.4	-9.4
ZINC02924334		0.406/4	DPM2n	-10.9	-10.5
NAC		N/A	—	-9.4	-10.5
NACD		N/A	—	-11.7	-13.2
NACDSW		N/A	—	-15.4	-13.2

NBP No realistic Binding Pose could be obtained

N/A Not applicable since no pharmacophore model was used in compound identification

Compounds ZINC02561129 and ZINC01683322 were selected from the drug-like subset of the ZINC database to be tested *in vitro* and to explore the DPM2 cavity (Table 3.15). Compound ZINC02561129 has a good best fit value of 3.309/4 with low docking energies (Table 3.15). Illustrated in Figure 3.38 are both its binding interactions and pose when docked to PfSpdSyn (PDBid 2I7C). The phenol ring of ZINC02561129 forms hydrogen bonds with Gln 72 and Glu 147, corresponding to similar interactions made by the sugar ring of dcAdoMet (Figure 3.38). Hydrogen bonds are also formed between ZINC02561129, Asp 121 and Asp 196, which are similar to the bonds formed between the aminopropyl group of dcAdoMet and PfSpdSyn. PfSpdSyn further makes a hydrogen bond with compound ZINC02561129 via Tyr 264. The aminobutyl chain of compound ZINC02561129 makes a hydrogen bond with Asp 199, which corresponds to the non-attacking nitrogen of putrescine and the amino group of 4MCHA (Figure 3.38).

Asp 199 is located on the gate-keeping loop of PfSpdSyn and it is anticipated that the hydrogen bond formed with it will stabilize the loop over the active site upon binding and allow the inhibitor to occupy the active site for longer. The described hydrogen bond network suggest that ZINC02561129 can be accommodated within the DPM2 binding cavity and potentially inhibit the enzyme.

Figure 3.38: Compound ZINC02561129 selected to explore the DPM2 binding cavity. *(Left)* Illustration of the docking pose ZINC02561129 assumes and its hydrogen bond formation with PfSpdSyn (PDBid 2I7C). *(Right)* ZINC02561129 overlaid with dcAdoMet and 4MCHA co-crystallized in PfSpdSyn (PDBid 2PT9).

Compound ZINC01683322 has a very low best fit value (0.285/4), but the docking scores were very good relative to the other docked compounds (Table 3.15). This compound contains a piperazine ring, representing the cyclohexylamine ring of the known inhibitor 4MCHA, with two aminopopyl chains sprouting from the nitrogens of the ring. It is possible that one of the aminopropyl chains can bind in the same binding pocket where the aminopropyl chain of dcAdoMet binds, while the other chain may bind within the non-attacking nitrogen binding pocket of putrescine (Figure 3.39). The aminopropyl chain's amino group binding in the dcAdoMet cavity forms hydrogen bonds with His 103, Asp 196 and Asp 127, which is similar to the aminopropyl chain of dcAdoMet. The other aminopropyl chain within the putrescine binding cavity forms hydrogen bonds with Asp 199 and Glu 231 to simulate the interaction by the non-attacking nitrogen of putrescine (Figure 3.39). The piperazine ring with its two nitrogens facilitates the bridging of the catalytic center, with the nitrogen closest to the dcAdoMet cavity forming a hydrogen bond with Ser 197 (Figure 3.39). It is thought that the hydrogen bonds

between ZINC01683322, Ser 197 and Asp 199 will bind the gate-keeping loop and allow the compound to occupy the active site longer and thus inhibits PfSpdSyn.
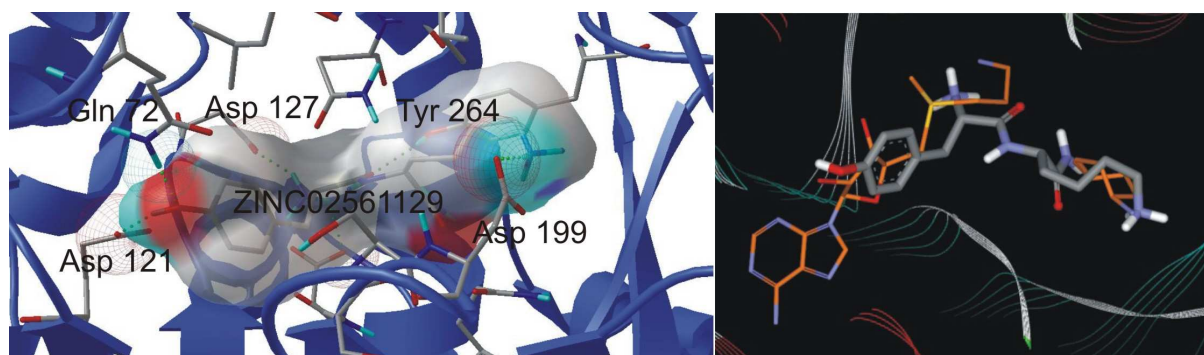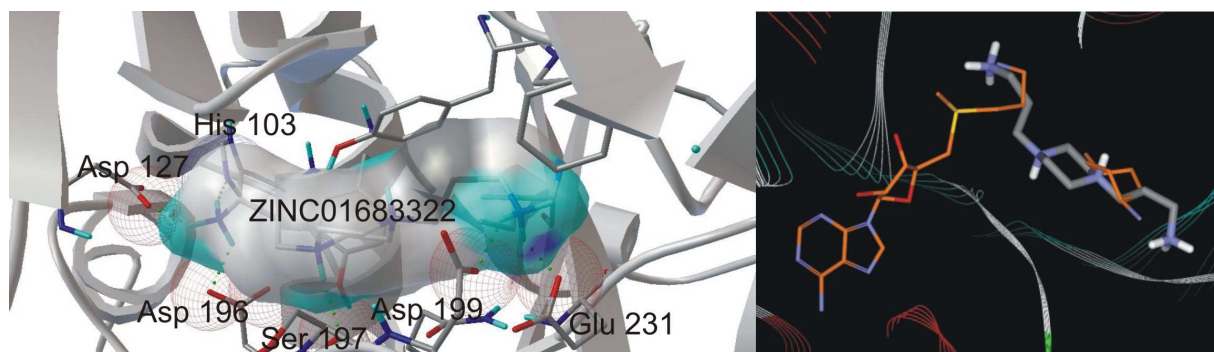

Figure 3.39: Compound ZINC01683322 selected to explore the DPM2 binding cavity. *(Left)* Illustration of the docking pose ZINC01683322 assumes and its the hydrogen bond formation with PfSpdSyn (PDBid 2I7C). *(Right)* ZINC01683322 overlaid with dcAdoMet and 4MCHA co-crystallized in PfSpdSyn (PDBid 2PT9).

A third compound 2-(3H-imidazol-4-yl)-N-methylethanamine, also known as histamine dihydrochloride (HDHC), was also selected to be tested on PfSpdSyn *in vitro*. HDHC was detected by performing a similarity search of imidazole rings using Scifinder and docked to PfSpdSyn with docking energies slightly better than 4MCHA. The reason for inclusion of this structure was that compounds containing an imidazole ring were frequently detected during the DPM screens against the drug-like subset of the ZINC database. The frequent detection of imidazole rings or derivatives thereof was also encountered in the study (2/28) by Jacobsson *et al.* (2008). Ten compounds contained a five membered aromatic ring resembling similar characteristics as the imidazole ring (Figure 3.14). From the study by Shirahata *et al.* (1991) it is known that benzene rings do not bind as well as cyclohexane rings and that 3-methylcyclopentylamine (compound 19; Table 3.2) does not bind as well as 4MCHA (compound 7; Table 3.2) in mammalian SpdSyn. Therefore, the inclusion of HDHC in *in vitro* tests allows exploration of the usefulness of imidazole rings as a substructure or scaffold in the design of new inhibitors against PfSpdSyn. Although not a definitive test to determine whether imidazole rings can be used or not, a better understanding will be obtained of inhibitors that bind within PfSpdSyn. Various docking poses were found for HDHC, the best scoring binding pose is shown in Figure 3.40. The

imidazole ring of HDHC forms hydrogen bonds with Tyr 103, Gln 93 and Ser 197. The ethylamine chain protruding from HDHC binds in the aminopropyl binding pocket of dcAdoMet, forming hydrogen bonds with Asp 127, His 103 and Asp 196. HDHC was also shown to bind within the putrescine binding cavity, however no good docking pose could be identified (results not shown).
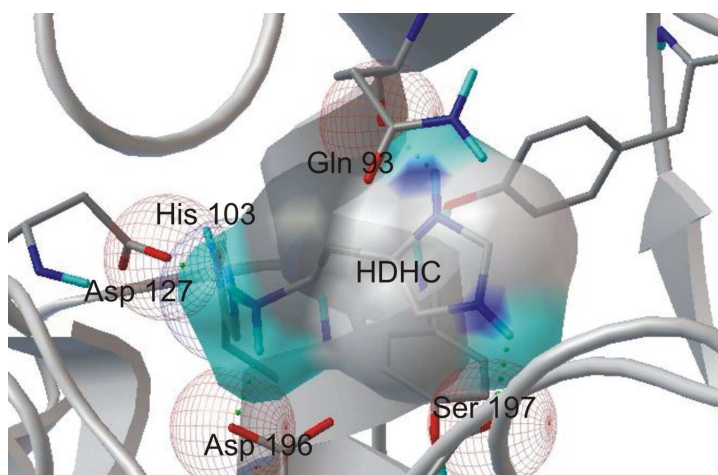
Figure 3.40: An illustration of the docking poses for HDHC within PfSpdSyn, PDBid 2I7C. In green are illustrated the hydrogens bonds formed between HDHC and residues Asp 127, His 103, Asp 196, Gln 93 and Ser 197.

The information obtained from MIF analysis of the DPM2 binding cavity was used together with PhFs confirmed by protein-ligand interactions to rationally derive possible inhibitors. The main idea behind the DPM2 binding cavity was to find a compound bridging from the putrescine binding cavity to the dcAdoMet cavity and thus crossing the catalytic center. The only inhibitor known to cross the highly negatively charged catalytic center is AdoDATO. The crossing of the catalytic center by an aliphatic chain can be attributed to the substrate-like characteristics of AdoDATO which are strong enough to allow this. It should be emphasized that 4MCHA and APE is six-times and five-times more potent inhibitors than AdoDATO, respectively (Haider et $al.$, 2005). Both 4MCHA and APE is much smaller molecules than AdoDATO and it is known that their strong inhibition is due to cooperative binding with either dcAdoMet or MTA. It can therefore be deduced that the higher K_i value of AdoDATO is due to the unfavourable interactions of the aliphatic part of the aminopropyl moiety which crosses the highly

negatively charged catalytic center. The information above highlights the importance of finding chemical entities which will be able to link inhibitors from the dcAdoMet and putrescine binding cavities by bridging the catalytic center.

It is known that the catalytic center binds positive ionisable groups to catalyze the transfer of an aminopropyl group. The residues involved in binding the attacking nitrogen of putrescine are Tyr 102, Asp 196 and the backbone carbonyl group of Ser 197 (Wu *et al.*, 2007). This is supported by the HsSpdSyn crystal structure co-crystallized with putrescine and MTA. MIF analysis of the chemical space surrounding the catalytic center was evaluated to derive properties, which could assist in finding better inhibitors. MIF analysis revealed both a HYD and PI PhF close to each other within the catalytic center (Figure 3.41). The HYD PhFs were detected in both the subensemble of structures sampled during the MD simulation and the PfSpdSyn monomers of PDBid 2PT9. The HYD PhF derived from the PfSpdSyn PDBid 2PT9 structures (-2.5 kcal/mol) has a stronger affinity than that derived from the subensemble of PfSpdSyn structures (-1.6 kcal/mol). This makes sense since the HYD PhF corresponds to the binding position of the methyl group of 4MHCA. The PI PhFs identified by MIF analysis of the subsensemble of structures sampled and that of the crystal structure of PDBid 2PT9 were found to be very similar in their binding energies, -20.2 and -20.8 kcal/mol, respectively. The PI PhF derived from the crystal structure PDBid 2PT9 is shifted almost 3Å from the PI PhF identified from the subensemble of structures sampled during the MD simulation (data not shown). This makes sense since PfSpdSyn would adjust itself to accommodate the methyl group within the active site/catalytic center. Therefore, the PI PhF derived from the subensemble of structures of PfSpdSyn was chosen to be used as the representative PhF.

The identified PhFs and the binding poses of AdoDATO, dcAdoMet and 4MCHA, which are known were used to derive compounds which can bridge the catalytic center. The compound N-(3-aminopropyl)-*trans*-cyclohexane-1,4-diamine (NACD) was rationally derived from the above considerations and represents a basic structure (scaffold)
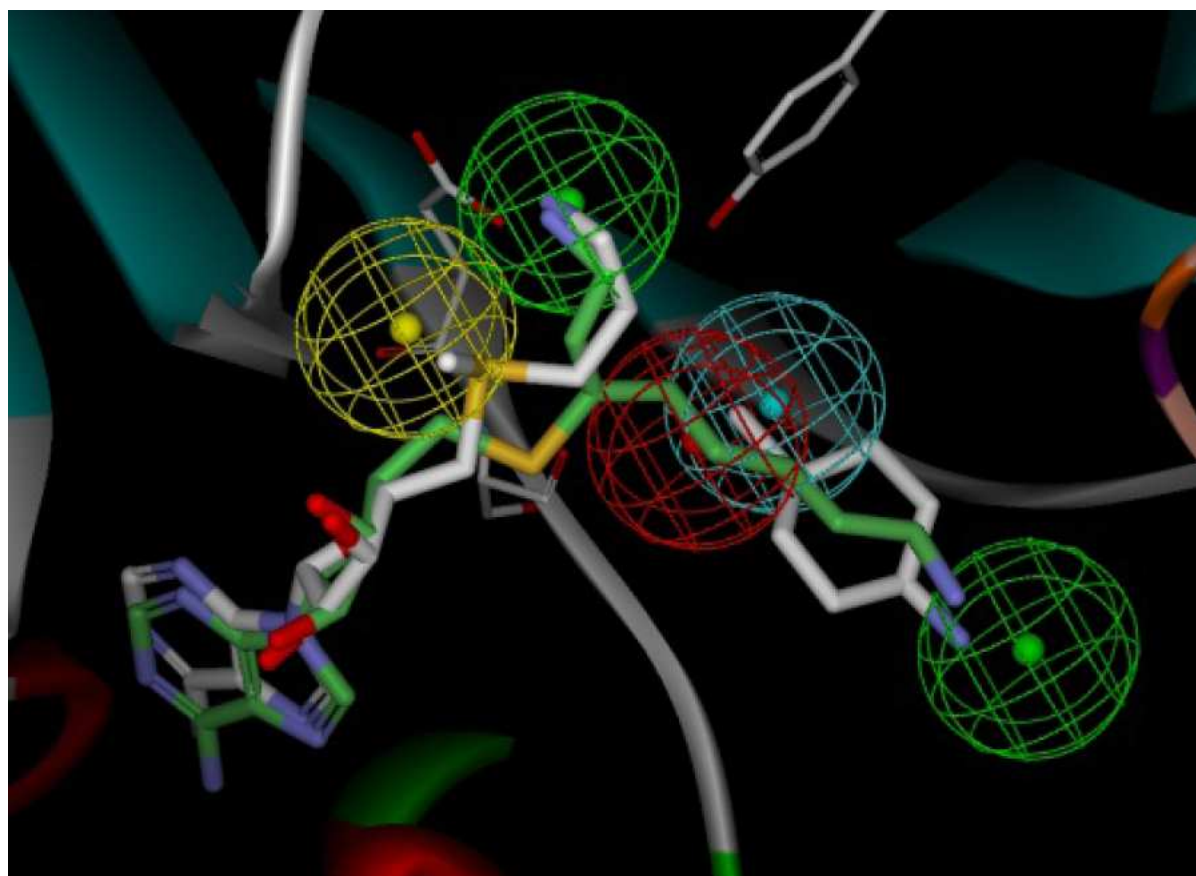
Figure 3.41: Binding hotspots identified in surrounding chemical space of the catalytic center of PfSpdSyn. AdoDATO from the PfSpdSyn (PDBid 2I7C) was overlaid with dcAdoMet and 4MCHA of PfSpdSyn (PDBid 2PT9) and is shown in stick representation. The residues represented are Tyr 102, Asp 196 and Ser 197. From MIF analysis a HYD PhF was identified near the methyl group of 4MCHA and is shown in blue (From PDBid 2PT9). A positive ionizable PhF was identified within the catalytic center from the subensemble of structures sampled during the MD simulation and is represented in red. The positive ionizable group is found within the chemical space where the attacking nitrogen of putrescine will bind. The yellow sphere reprsents a PI sphere near to the positively charged sulpher of dcAdoMet. The green spheres show the PI PhFs identified to represent the nitrogen of the aminopropyl chain from dcAdoMet and the non-attacking nitrogen of putrescine (amino group of 4MCHA).

for an inhibitor of PfSpdSyn. NACD is not commercially available, but it has been tested for inhibition against deoxyhypusine synthase by Jakus *et al.* (1993) and found not to inhibit the latter. NACD was docked to both the PfSpdSyn structures (PDBid 2PT9 and 2I7C) and the expected binding poses with good binding energies were obtained (Table 3.15). The cyclohexylamine moiety of NACD binds in a similar manner to 4MCHA and the corresponding amino group forms a hydrogen bond with Asp 199 suggesting similar interactions as 4MCHA with PfSpdSyn (Figure 3.42). The bridging amino group, the

nitrogen connecting the aminopropyl chain of NACD to the cyclohexyl ring (the nitrogen closest to the ring), forms two hydrogen bonds, one with Tyr 102 and another with the backbone carbonyl group of Ser 197. It is expected that these hydrogen bonds will reduce the binding penalty an aliphatic carbon would have bridging the catalytic center and thus increase the binding affinity and inhibition. The aminopropyl chain of NACD binds to the same cavity as the aminopropyl chain of dcAdoMet and forms hydrogen bonds with His 103, Asp 127 and Asp 196, respectively (Figure 3.42).
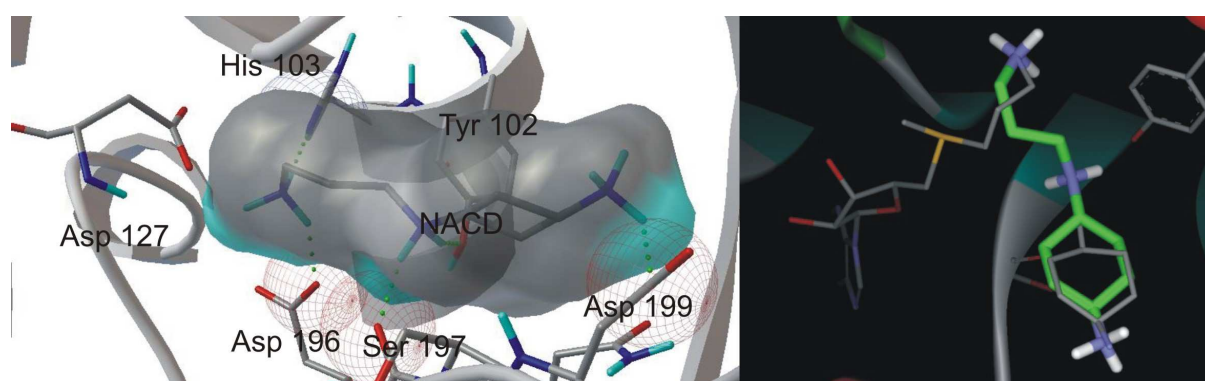
Figure 3.42: An illustration of NACD docked to PfSpdSyn. *(Left)* NACD docked in the expected binding pose forming hydrogen bonds with Tyr 102, His 103, Asp 196, Ser 197 and Asp 199. *(Right)* An overlay of NACD (green) with dcAdoMet and 4MHCA in the PfSpdSyn (PDBid 2PT9) active site.

SciFinder searches were performed to identify similar compounds to NACD. N-(3-aminopropyl)-cyclohexylamine (NAC) was identified as a similar and affordable compound and was docked to PfSpdSyn to evaluate its binding pose and docking energies. Good binding poses and low binding energies were obtained (Table 3.15; Figure 3.43). NAC differs from NACD in that its ring moiety is a cyclohexylamine and not a 1,4-diaminocyclohexyl ring and therefore takes up the same binding pose as NACD with the same hydrogen bond pattern except for the missing amino group (Table 3.15). This makes NAC a good second choice to test as a bridge between the dcAdoMet and putrescine binding cavities and the effect thereof on inhibition.

Further investigation of the DPM2 binding cavity revealed a PI PhF with a very strong binding affinity (-23.2 and -28.4 kcal/mol for the subensemble of structures and PfSpdSyn,

Figure 3.43: An illustration of the docking pose NAC assumes when docked to PfSpdSyn. *(Left)* NAC docked to PfSpdSyn (PDBid 2PT9). Hydrogen bond formation can be seen between the aminopropyl group of NAC and Asp 127, His 103 and Asp 196. The bridging nitrogen has hydrogen bonds with Tyr 102 and the carbonyl backbone oxygen of Ser 197. *(Right)* NAC overlaid with dcAdoMet and 4MCHA in PfSpdSyn (PDBid 2PT9).

PDBid 2PT9, respectively) within the chemical space that the positively charged sulphur of dcAdoMet normally occupies when bound to PfSpdSyn (Figure 3.41). To further explor an additional ethylamine group was added to the aminopropyl chain at position number two. This compound will be referred to as NACDSW (Figure 3.44). NACDSW docked to PfSpdSyn with similar binding poses to those of NAC and NACD. The additional aminoethyl group protrudes into the chemical space of the new PI PhF discussed above. The docking energies obtained were found to be significantly lower than that of NAC and NACD (Table 3.15). The newly-added aminoethyl group was shown to form hydrogen bonds with Asp 127 and Asp 196 (Figure 3.44). In Figure 3.44 the hydrogen bonds formed by the bridging nitrogen is not shown but it should be noted that with optimization of the structure they are indeed present.

In vitro data for both compounds ZINC02561129 and ZINC01683322 showed no inhibition at a 100 μM concentration. As discussed under the DPM1 binding cavity section the putrescine binding cavity is thought to accommodate an aminopropyl chain sprouting from the piperidine ring and this was supported by the *in vitro* testing of compound ZINC02582060. For compound ZINC01683322, it was thought that the piperazine ring and one of the aminopropyl chains would be accommodated within the putrescine binding cavity in a similar manner to compound ZINC02582060 (Figure 3.32). If this is so
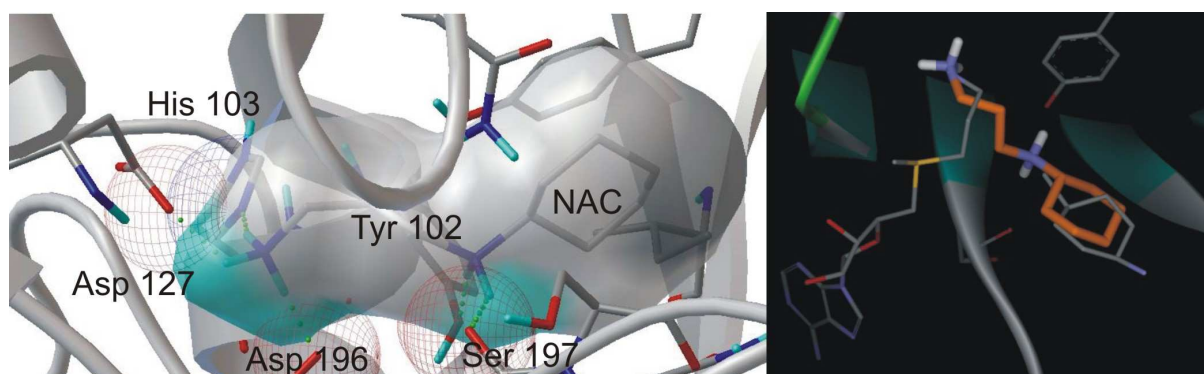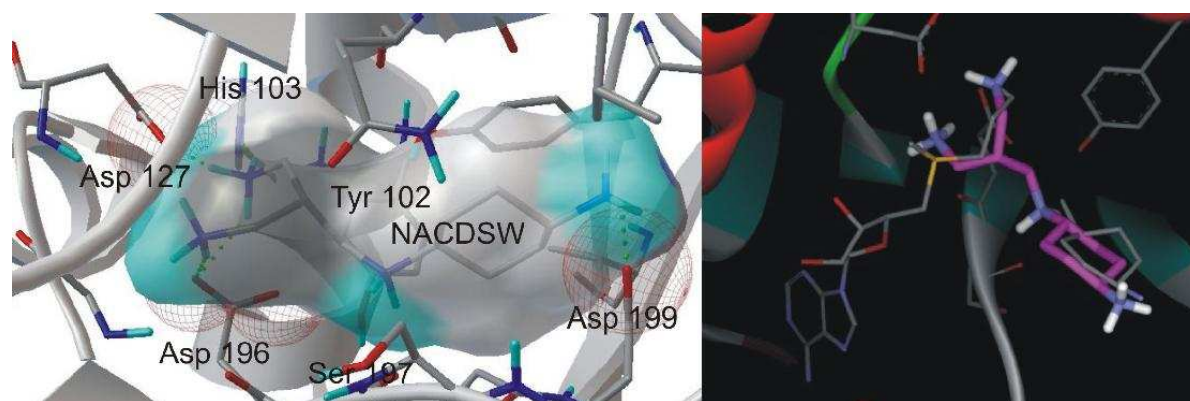
Figure 3.44: An illustration of the docking pose NACDSW assumes when docked to PfSpdSyn. *(Left)* NACDSW docked to PfSpdSyn forming hydrogen bonds (green dotted line) with Asp 127, Asp 196, His103, Tyr 102 and Asp 199. *(Right)* NACDSW (purple) overlaid with dcAdoMet and 4MCHA of PfSpdSyn (PDBid 2PT9).

the second aminopropyl chain sprouting from the piperazine will be able to bridge the catalytic center and bind in the dcAdoMet binding cavity. The lack of inhibition by compound ZINC01683322 can probably be explained by considering the information gained after *in vitro* testing of compound 4AMP. This compound was selected for exploration of the DPM1 binding cavity to determine the effect of the ring nitrogen of the piperidine ring when positioned closest to the catalytic center (DPM1 binding cavity section). It was evident that the piperidine nitrogen could not be accommodated in the proposed chemical space within the putrescine binding cavity. If compound ZINC01683322 binds to the DPM2 binding cavity as was proposed by the docking studies, its piperazine ring nitrogens would occupy this disallowed chemical space and prevent its binding as was shown with the *in vitro* results of 4AMP. This then explains the lack of inhibition of compound ZINC01683322. HDHC was selected to test whether an imidazole ring could be accommodated within the PfSpdSyn binding cavity. No significant reduction in PfSpdSyn activity could be observed in the presence of 100 μM HDHC. This however is not conclusive proof that imidazole rings are not accommodated within the PfSpdSyn binding cavity.

In vitro testing of the compound NAC at a 100 μM concentration showed a remarkable 86% reduction in the PfSpdSyn activity, which warranted the further investigation of this
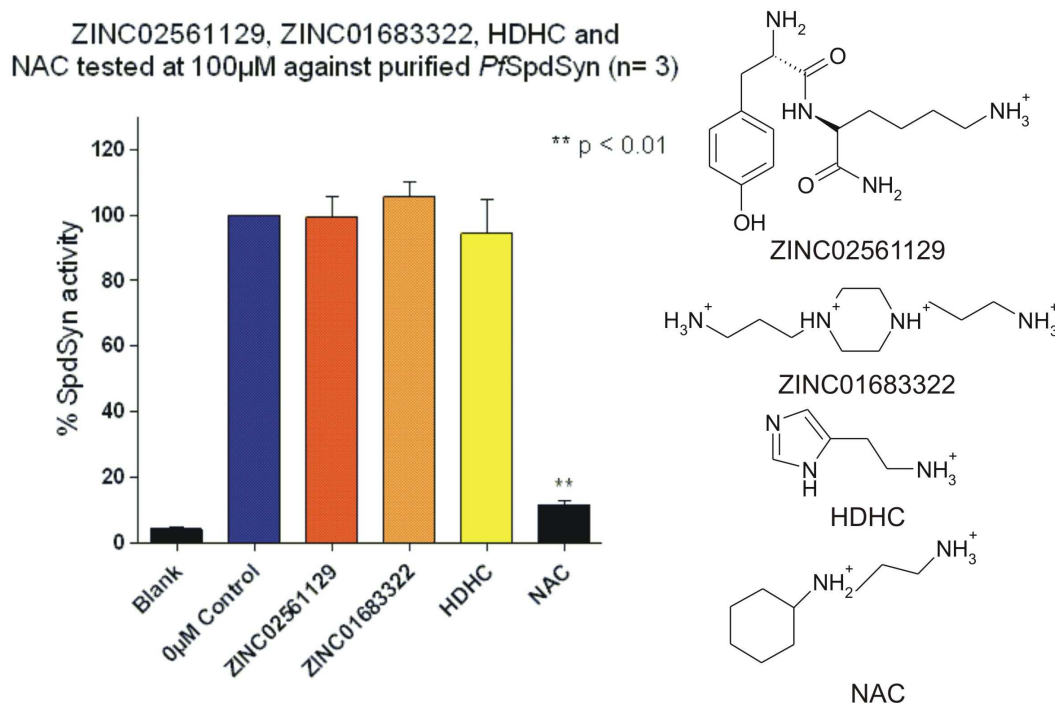
Figure 3.45: Inhibitory effect of 100 μM of ZINC02561129, ZINC01683322, HDHC and NAC when tested against PfSpdSyn. *(Left)* A bar graph depicting the reduction in PfSpdSyn activity of ZINC02561129, ZINC01683322, HDHC and NAC at a 100 μM concentration. *(Right)* The chemical structures of the compounds tested *in vitro*.

compound (Figure 3.45). Enzyme kinetics analysis was subsequently performed for the compound and the K_i value determined (Figure 3.46).

NAC inhibitory activity was determined at three different concentrations of putrescine. The dcAdoMet concentration was kept constant at 100 μM. Data from a Lineweaver-Burk extrapolation indicates a similar K_m value for putrescine at 25.1 μM and slightly lowered V_{max} values at 96.4 μmol/min/mg than previously reported by Haider *et al.* (2005) (Figure 3.46). The lowered V_{max} could possibly be attributed to the alternative source of dcAdoMet used in this study (R/S isomers of dcAdoMet; Ingrid Muller, personal communication). Additionally, in the presence of compound NAC the K_m and V_{max} parameters of PfSpdSyn were affected (Figure 3.46). For compound NAC to be a true competitive inhibitor of the putrescine binding site, only the K_m value is expected to change. However, the kinetic data shows that the V_{max} was not re-established at putrescine concentrations far greater than its K_m. This suggests that compound NAC also affects the binding of

dcAdoMet and that the additional putrescine is not able to disrupt the tight binding interaction. A secondary plot from the Lineweaver-Burk plot was used to calculate the K_i value of compound NAC and it was found to be 2.8μM. The *in vitro* studies were performed by S Reeksting.

The low K_i value of compound NAC is probably due to the hydrogen bonds which it forms with PfSpdSyn (Figure 3.43). This, however, can only be true if the compound NAC binds in the proposed docking pose (Figure 3.43), which is supported by the kinetic data. It can be deduced that compound NAC does not only compete with putrescine but also with dcAdoMet. This would make sense since the aminopropyl chain of the compound NAC supposedly binds in the aminopropyl binding pocket of the dcAdoMet binding cavity. The *in vitro* data of the compounds tested for the DPM1 binding cavity, however, suggested that the putrescine binding cavity will not be able to accommodate the aminopropyl chain of the compound NAC. This observation is supported by comparison of the loss in the activity of PfSpdSyn at 100 μM concentrations of compounds NAC and ZINC02582060. Compound ZINC02582060 only reduced the activity by 18% whereas the compound NAC showed a reduction of 86%. Thus it is thought to be unlikely that the larger of the two compounds, compound NAC, would be accommodated within the putrescine binding cavity. It can therefore be said with a high degree of confidence that the compound NAC binds to PfSpdSyn in the way that was proposed (Figure 3.43).

The low K_i value of compound NAC is therefore suggested to be mainly due to the ability of the compound to bridge the catalytic center and that the hydrogen bonds between the compound NAC and residues Tyr 102 and Ser 197 significantly contribute to the strong binding of the compound in the active site of PfSpdSyn. These hydrogen bonds are proposed to play a role in the stabilization of the gate-keeping loop upon binding and to keep it closed for a longer period over the active site.

Compounds such as NACD and NACDSW will probably have much higher binding affinities than compound NAC, since it will also bind to the gate-keeping loop at the

non-attacking end of the putrescine binding cavity. These compounds are also expected to have much lower K_i values.

	0µM Inhibitor	20µM Inhibitor	40µM Inhibitor
Slope	0.2891 ± 0.02640	3.716 ± 0.2692	6.321 ± 0.2597
Y-intercept when X=0.0	0.01027 ± 0.0006868	0.008451 ± 0.007002	0.001158 ± 0.006756
X-intercept when Y=0.0	-0.03554	-0.002274	-0.0001832
1/slope	3.459	0.2691	0.1582

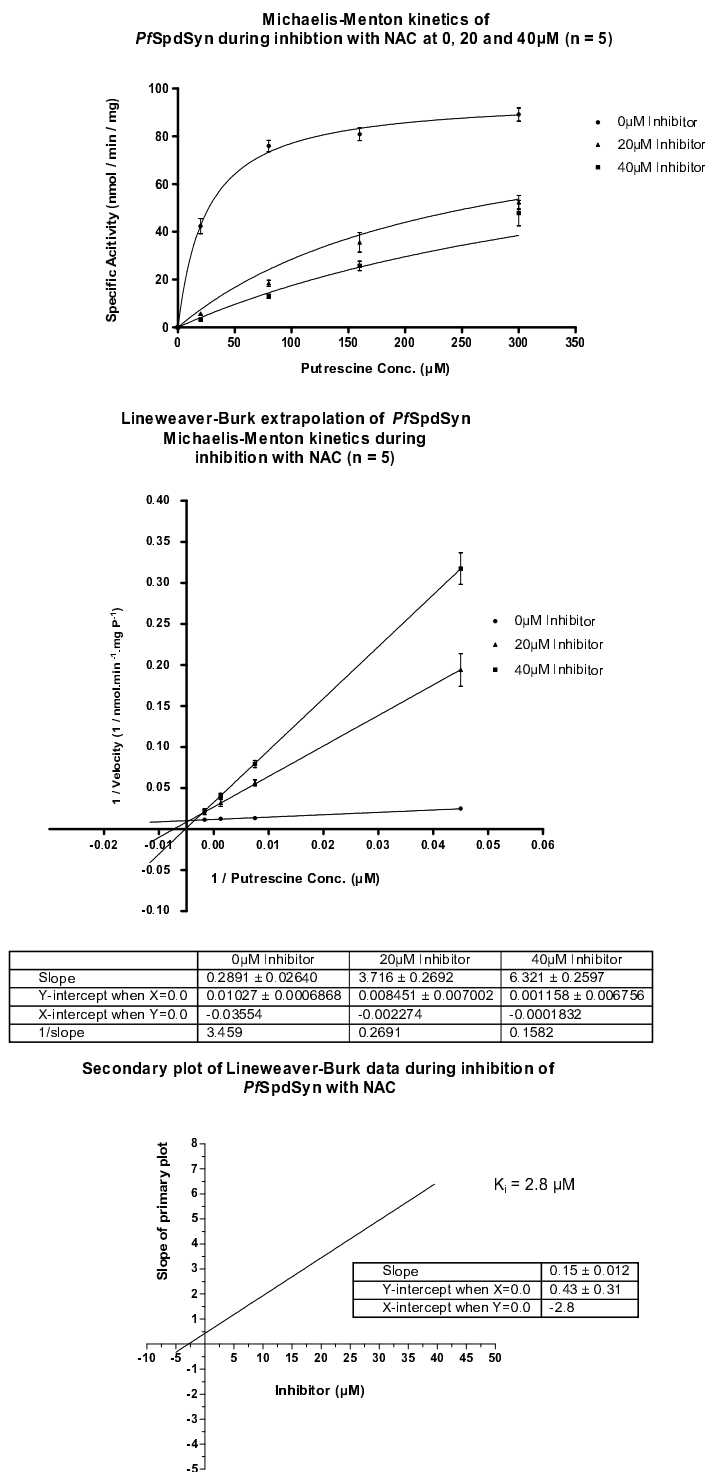| Slope	0.15 ± 0.012
Y-intercept when X=0.0	0.43 ± 0.31
X-intercept when Y=0.0	-2.8

Figure 3.46: Kinetic data determined for compound NAC tested against PfSpdSyn.

3.3.3.6. DPM3 Binding Cavity

The DPM3 binding cavity was specifically selected to explore the dcAdoMet binding cavity and include the catalytic center (Figure 3.28). Less focus was placed on the DPM3 binding cavity than with the DPM1 and DPM2 binding cavities, since it binds adenosyl which is a widely used substrate moiety in living organisms. Inhibitors against this cavity may thus bind non-specifically and have various other effects due to its binding to unrelated targets. The PhFs derived from MIF analysis were used to generate a shortlist of PhFs best describing the DPM3 binding cavity. Before performing database screens the PhFs were tested to see if it was possible to identify dcAdoMet and was demonstrated to have a best fit value of 0.899/4 (Table 3.17; Figure 3.47). Docking of dcAdoMet to both the PfSpdSyn structures yielded a docking energy of -13.2 kcal/mol when docked to PfSpdSyn (PDBid 2I7C) with a similar docking pose as the crystallized pose of PfSpdSyn (PDBid 2PT9) revealed by overlaying (data not shown). When docked to PfSpdSyn PDBid 2PT9, co-crystallized with dcAdoMet and 4MCHA, no pose representing the crystallized dcAdoMet could be illustrated. Therefore, during the selection of compounds to be tested *in vitro,* a compound was considered to carry more weight when it could be docked to PfSpdSyn (PDBid 2I7C).
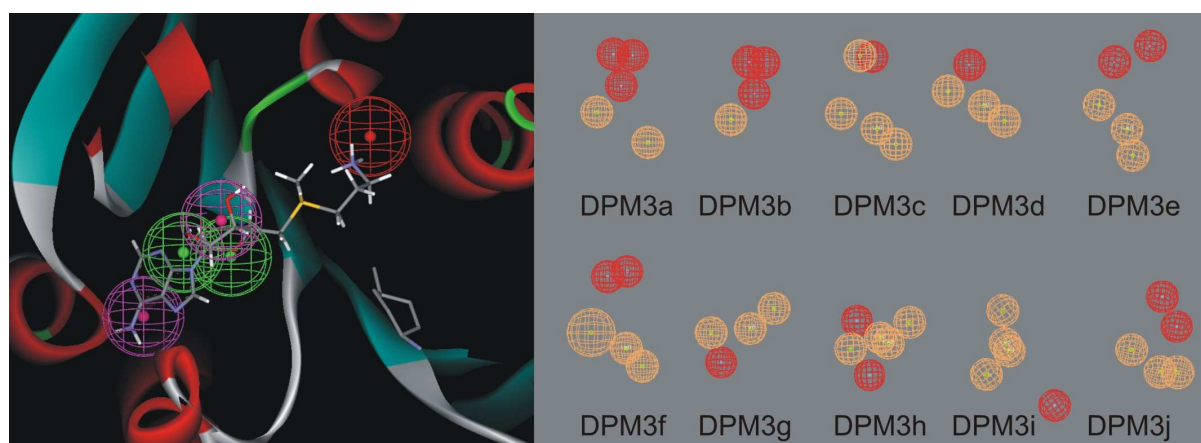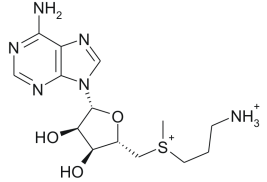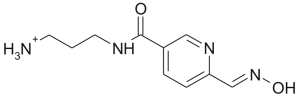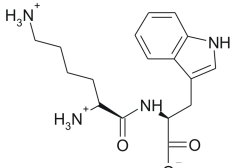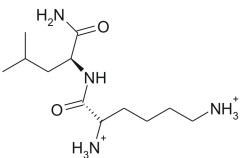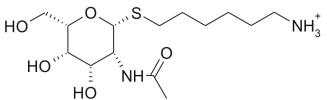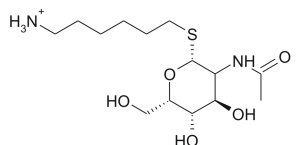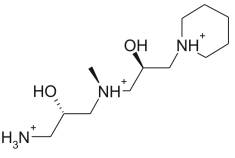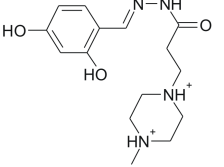
Figure 3.47: An illustration of dcAdoMet within PfSpdSyn and the DPMs used to explore the DPM3 binding cavity. *(Left)* An overlay of the PhFs used to identify dcAdoMet within PfSpdSyn. Illustrated in a red sphere is a positive ionizable group, in purple hydrogen bond donors and in green hydrogen bond acceptors. *(Right)* The DPMs selected to explore the chemical space of the DPM3 binding cavity. Displayed in red are positive ionizable groups and in orange either hydrogen bond donors or acceptors.

Table 3.17: Compounds used in the exploration of the DPM3 binding region. These compounds were identified from DPM screens against the drug-like subset of the ZINC database. The DPMs are representative of the chemical characteristics found within the DPM3 binding region.

Compound id	Structure	Best Fit Value/ Number of PhFs	DPM model	Docking Energies	
				2I7C	2PT9
dcAdoMet		0.899/4	DPM3g	-13.2	NBP
ZINC01901746		0.293/4	DPM3g	-10.0	-8.8
ZINC02384979		0.496/5	DPM3c	-11.5	-9.9
ZINC02556625		0.172/4	DPM3g	-11.8	-10.4
ZINC04533981		0.823/4	DPM3g	NBP	NBP
ZINC04533983		0.833/4	DPM3g	-10.1	-7.4
ZINC04955838		0.506/5	DPM3f	-13.2	-9.8
ZINC05030866		0.727/4	DPM3g	-12.0	-10.5

NBP No realistic Binding Pose could be obtained

A total of ten DPMs (DPM3a-DPM3j) were constructed consisting of between four and six PhFs (Figure 3.47). These DPMs consisted of PhFs from both the sub-ensemble

of PfSpdSyn structures and the PfSpdSyn crystal structure, PDBid 2PT9. The DPMs were screened against the drug-like subset of the ZINC database and resulted in between 0 and 1813 hits. The best fit values of all the compounds were calculated and the hits were ranked accordingly. The compounds were visually inspected to select the best fitting candidates and further analyzed by performing docking studies. Eight compounds were selected to be docked to both the PfSpdSyn structures used in docking (Table 3.17).

Compounds ZINC04955838 and ZINC05030866 were selected to be tested for inhibitory activity against PfSpdSyn *in vitro* (Table 3.17). Compound ZINC04955838 was docked to PfSpdSyn and yielded a best fit value of 0.506/5 and good docking energies (Table 3.17; Figure 3.48). Evaluation of the binding poses of compound ZINC04955838 delivered various possible docking poses with significant variation between them. The docking pose of ZINC04955838 displayed in Figure 3.48 fit the dcAdoMet cavity of which the best was selected by visual inspection. The docking energy of compound ZINC04955838 against PfSpdSyn (PDBid 2I7C) was marginally lower than that of dcAdoMet and is suggestive of strong binding. Compound ZINC04955838 forms hydrogen bonds with Gly 124, Asp 127 and Asp 196 of PfSpdSyn, which indicate that these hydrogen bonds will allow compound ZINC04955838 to bind within the dcAdoMet binding cavity. There is also the possibility of co-operative binding in the presence of putrescine.
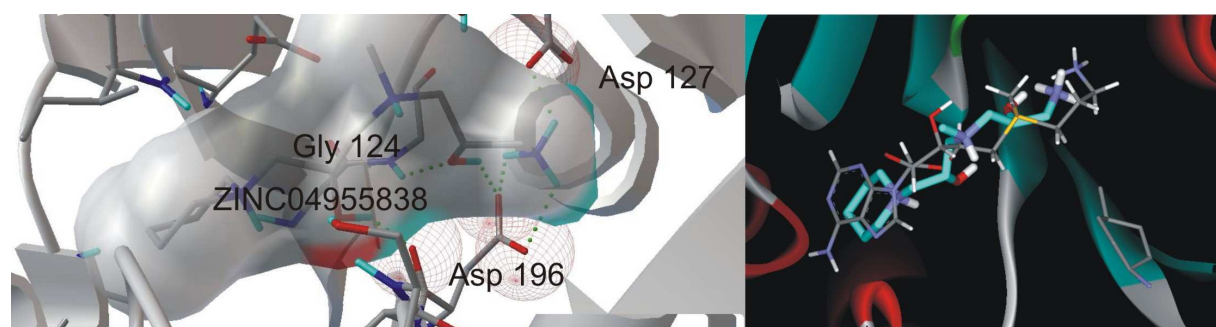
Figure 3.48: ZINC04955838 docked to PfSpdSyn. *(Left)* The docking pose of ZINC04955838 docked to PfSpdSyn. Illustrated with green dots are hydrogen bonds formed between ZINC04955838 and the Gly 124, Asp 127 and Asp 196 residues of PfSpdSyn. *(Right)* ZINC04955838 overlaid with dcAdoMet in the PfSpdSyn (PDBid 2PT9) active site.

ZINC05030866 was selected to explore the possibility that a phenol group can bind within the chemical space of the adenosyl moiety of dcAdoMet. Docking studies showed a fit value of 0.727/4 and good docking energies (Table 3.17; Figure 3.49). The characteristics of compounds ZINC05030866 and ZINC04955838 are not the most desirable (e.g. low best fit value), however they were the compounds found to best fit the DPM3 binding cavity. The drug-like subset of the ZINC database was also searched for substructures containing the adenosyl moiety but none were found. No further similarity searches were performed on dcAdoMet.
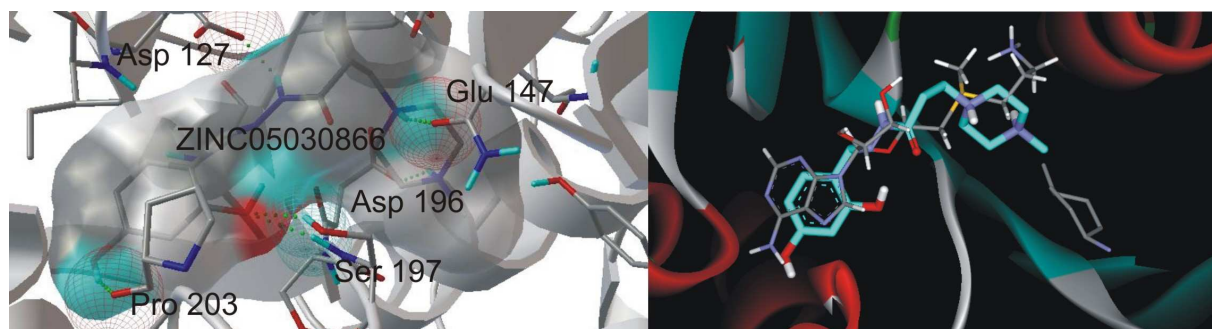
Figure 3.49: An illustration of the docking pose of ZINC05030866 docked to PfSpdSyn. *(Left)* ZINC05030866 docked to PfSpdSyn (PDBid 2I7C). Green dots represent the hydrogen bonds formed between ZINC05030866 and PfSpdSyn these bonds are made with residues Pro 203, Asp 196, Ser 197, Glu 147 and Gln 93. *(Right)* ZINC05030866 overlaid with dcAdoMet in the dcAdoMet binding cavity of PfSpdSyn (PDBid 2PT9).

Both compounds ZINC04955838 and ZINC05030866 were purchased and tested *in vitro*. Neither compound showed any reduction in the PfSpdSyn activity at a 100 μM concentration. Compound ZINC04955838 was selected in an attempt to identify a new scaffold which may bind to the dcAdoMet cavity. The purpose for selection of compound ZINC05030866 was two-fold. Firstly, to identify if this scaffold can bind the dcAdoMet cavity. Secondly, to explore the possibility that the phenol and piperazine moieties can be accommodated in their respective proposed binding cavities. Both of these hypotheses were shown not to be true, even though they were the best scoring compounds identified through pharmacophore screening and docking studies of the dcAdoMet cavity.

3.3.3.7. DPM4 Binding Cavity

The DPM4 binding cavity was the last of the four cavities being explored to find new inhibitors for PfSpdSyn and included the whole of the active site. The DPM4 binding cavity was expected to result in only a few compounds since it is a big cavity with PhFs being as far apart as 14Å. For example, the PhF representing the N6 of the adenosyl moiety and the PhF representing the nitrogen of the aminopropyl group are 14Å apart, whereas the latter is also 14Å apart from the nitrogen of the amoinopentyl group of AdoDATO. Therefore, few compounds having these PhFs are expected to be in the drug-like subset of the ZINC database. AdoDATO is the only known inhibitor that occupies the whole of the PfSpdSyn active site. The interactions AdoDATO makes with PfSpdSyn were discussed in section 3.1.3 and illustrated in Figure 3.12. As with DPM1-3 a shortlist of PhFs were made, which best described the DPM4 binding cavity and was used to construct eight DPMs (DPM4a-DPM4h; Figure 3.50). Before the drug-like subset of the ZINC database was screened the DPMs were used to see if they could be used to identify AdoDATO. AdoDATO had a best fit value of 0.796/5 (Table 3.18) which is illustrated in Figure 3.50.
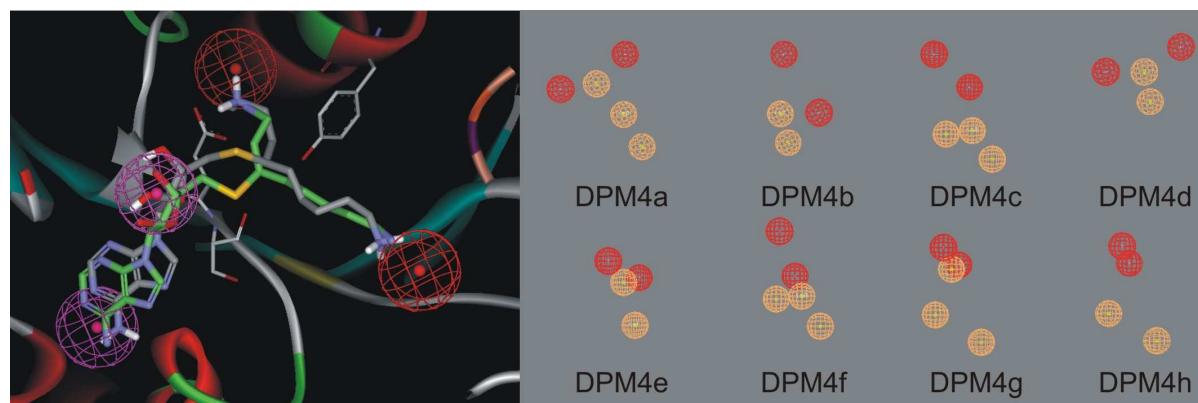
Figure 3.50: Illustration of AdoDATO within the active site of PfSpdSyn and the DPMs used to explore the DPM4 binding cavity. *(Left)* AdoDATO co-crystallized with PfSpdSyn (PDBid 2I7C) is displayed in green sticks. AdoDATO docked to PfSpdSyn (PDBid 2I7C) is displayed in gray sticks. The red mesh spheres represent positive ionizable groups and the purple spheres represent hydrogen bond donors. *(Right)* The DPMs used to explore the DPM4 binding cavity. In red are represented positive ionizable groups and in orange both the hydrogen bond donors and acceptors.
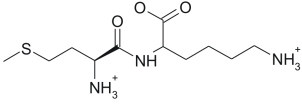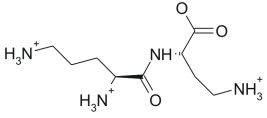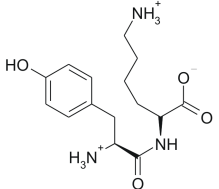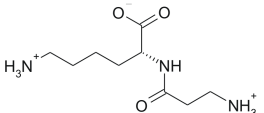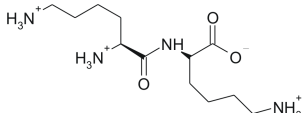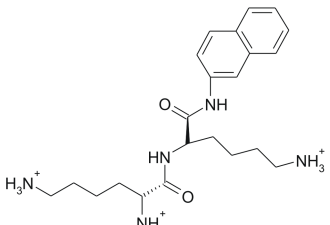
AdoDATO was docked back to PfSpdSyn (PDBid 2I7C) with a docking energy -17.1

kcal/mol and having the same orientation (docking pose; RMSD 1.05Å) as the crystallized AdoDATO (Figure 3.50). However, as mentioned before when AdoDATO was docked to PfSpdSyn, PDBid 2PT9 no docking pose could be demonstrated representing a similar orientation as the one in which AdoDATO was crystallized.

The eight DPMs constructed for the DPM4 binding cavity consisted of between 4 and 5 PhFs, which were selected from both the subensemble of structures sampled from the MD simulation and the crystal structure (PDBid 2PT9) used during MIF analysis. These DPMs were searched against the drug-like subset of the ZINC database and resulted in between 0 and 80 hits. The best fit values were calculated for all the compounds identified and all were visually inspected. Eight compounds were selected to be docked against PfSpdSyn with only ZINC04899780 occupying the whole active site (Table 3.18).

ZINC04899780 had a best fit value of 0.649/5 and was shown to have a docking energy of -11.6 kcal/mol to PfSpdSyn, PDBid 2I7C (Figure 3.51; Table 3.18). No realistic docking poses were obtained after docking ZINC04899780 to PfSpdSyn, PDBid 2PT9, but this result was not surprising since AdoDATO could also not be docked to this PfSpdSyn. Compound ZINC04899780 forms hydrogen bonds with residues Asp 127 and Gln 93. Figure 3.51 demonstrates the location of two nitrogens from compound ZINC04899780 within the binding pockets of the nitrogens of both the aminopropyl and aminopentyl moieties of AdoDATO. A third nitrogen is found within the vicinity of the identified PI PhF (Figure 3.41) where the positive sulphur of dcAdoMet binds (Figure 3.51). The nitrogens in these binding pockets have the potential to rearrange themselves in order to form a hydrogen bond network, which will increase the binding affinity and possibly inhibition of the enzyme. Although compound ZINC04899780 shows less than the desirable properties (e.g. low best fit value) it was selected since it was the only compound demonstrated to occupy the DPM4 binding cavity, and thus the biggest part of the active site (Figure 3.51). It was expected that if compound ZINC04899780 was to show inhibition, it would be only at a high concentration. Since no compounds containing the adenosyl moiety are included within the drug-like subset of the ZINC database as

Table 3.18: Compounds used in the exploration of the DPM4 binding region. These compounds were identified from DPM screens against the drug-like subset of the ZINC database. The DPMs are representative of the chemical characteristics found within the DPM4 binding region.

Compound id	Structure	Best Fit Value/ Number of PhFs	DPM model	Docking Energies PDBid 2I7C	PDBid 2PT9
ZINC02522671		0.234/5	DPM4f	NBP	-9.0
ZINC02561080		0.465/5	DPM4f	-11.4	-9.9
ZINC02575131		0.476/5	DPM4f	NBP	-10.5
ZINC04533949		0.143/5	DPM4f	NBP	NBP
ZINC04556738		0.648/5	DPM4f	-11.6	-9.5
ZINC04899780		0.649/5	DPM4f	-11.6	NBP

NBP No realistic binding pose could be obtained

described in the previous section (DPM3), no similarity searches were performed. This skepticism of compound ZINC04899780 was confirmed by *in vitro* testing where it did not show any inhibition at a 100 μM concentration.

3.4. Conclusion

It is clear from the quality assessment analysis that the crystal structures used in this study, PDBid 2I7C and 2PT9, are of high quality and thus inspire confidence in the
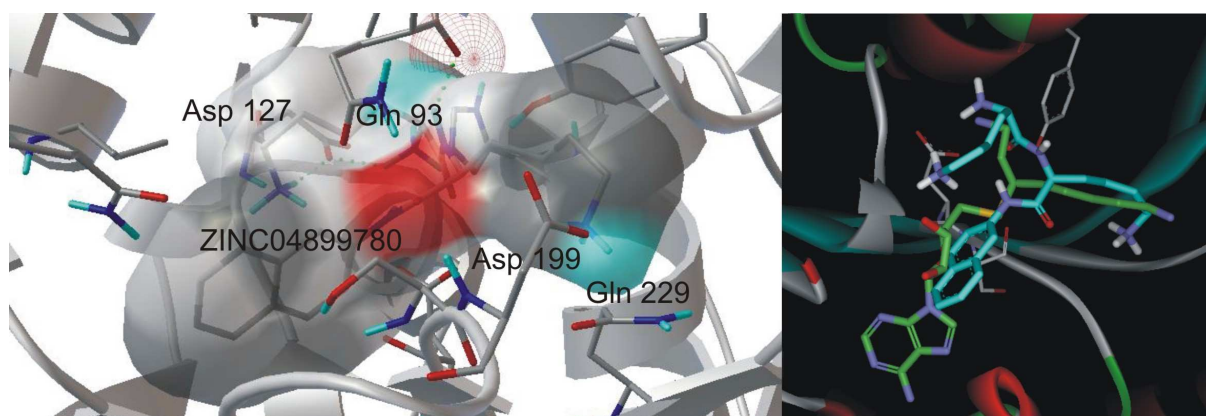
Figure 3.51: An illustration of compound ZINC04899780 docked to PfSpdSyn (PDBid 2I7C). *(Left)* Compound ZINC04899780 forms hydrogen bond with Asp 127, Gln 93 and Gln 229 (green dots). *(Right)* An overlay of compound ZINC04899780 with AdoDATO within the PfSpdSyn active site.

conclusions drawn from them. A MD simulation was performed on PfSpdSyn (PDBid 2I7C) and the trajectory from the production run was clustered to find the most representative structures of the sampled phase space. Six structures were identified and subsequently used in further analysis. The three monomers of PfSpdSyn (PDBid 2PT9) co-crystallized with 4MCHA and dcAdoMet were also analyzed and included in further analysis, in order to get a better representation of the putrescine binding cavity. Negative image construction was performed using MIF analysis on these nine structures from which PhFs were identified. The PfSpdSyn active site was further subdivided into four binding regions, DPM1 to DPM4. For each of the binding cavities DPMs were generated and searched against the drug-like subset of the ZINC database. Nine compounds all-together were identified as possible inhibitors and tested *in vitro*.

Two compounds, ZINC02582060 and 4AMP, were selected to explore the DPM1 binding cavity and found to reduce the PfSpdSyn activity by 18% and 5%, respectively. From the *in vitro* testing of these compounds it was concluded that PfSpdSyn can accommodate a piperidine ring within the putrescine binding cavity, only if the ring nitrogen is oriented towards the catalytic center. Four compounds were selected and tested *in vitro* to explore the DPM2 binding cavity. This binding region was selected to identify compounds, which could bridge the catalytic center by binding in both the putrescine and dcAdoMet bind-

ing cavities. Three of the selected compounds ZINC02561129, ZINC01683322 and HDHC showed no significant reduction in the PfSpdSyn activity. The fourth compound NAC showed a reduction of 86% of the PfSpdSyn activity at 100 μM and a K_i value of $2.8\mu M$. It was also confirmed by kinetic studies that NAC affects the binding of dcAdoMet which supports the binding mode to PfSpdSyn as proposed by the docking studies. The compound NAC is a novel hit compound for PfSpdSyn. Two compounds, ZINC04955338 and ZINC05030866, were selected to identify new scaffolds which would compete with dcAdoMet for binding in the DPM3 binding cavity. However, neither of these compounds showed any reduction of the PfSpdSyn activity when tested. ZINC04899780 was selected to bind within the DPM4 binding cavity, however no reduction in the PfSpdSyn activity was found.

Information gained from MIF analysis was also used to identify novel PhFs, which can assist in the optimization of lead compounds. From this a compound 2AMCHA was rationally derived and proven worthy of testing *in vitro* to verify this newly-identified binding hotspot. If this PhF can be verified it can be used in the optimization of 4MCHA and the newly-identified hit compound NAC. Furthermore, two structurally-similar compounds NACD and NACDSW have been proposed as possible inhibitors against PfSpdSyn, since stronger binding affinities are expected in comparison to the compound NAC. From the inhibition studies of compound NAC it can be deduced that the bridging of the catalytic center by the cyclohexylamine nitrogen play a key role in the inhibition of PfSpdSyn. This new binding hotspot can be exploited during lead optimization by the use of compounds NACD and NACDSW.

A comparison between the current study and the pharmacophore study by Jacobsson *et al.* (2008) revealed remarkable similarities between the compound structures identified during the respective database screens. Jacobsson *et al.* (2008) identified seven compounds out of twenty eight to be reversible binders using NMR experiments (no kinetic analysis were performed). The strongest binding compound (number 4) was suggested to have an IC_{50} value similar to MTA ($159\mu M$), and can therefore not be considered as

a lead. Incorporation of knowledge from various stages of the lead identification process is pivotal in structure-based drug design as shown in this study.

It can therefore be concluded that a novel hit compound NAC has been identified for PfSpdSyn. Various chemical modifications of the NAC scaffold have been suggested from knowledge gained during the development of the dynamic pharmacophore model which should be used during lead optimization.

Chapter 4

Concluding Discussion

Malaria affects the daily lives of more than 2 billion people worldwide and has been estimated to result in 300-500 million clinical cases annually leading to approximately 2 million deaths (Cambell *et al.*, 2004; Guerra *et al.*, 2008). These deaths are mainly caused by the most virulent malaria species, *P. falciparum* (Cambell *et al.*, 2004; Snow *et al.*, 2005). An underestimation of the malaria burden has recently been reported, showing the global distribution of clinical episodes of only *P. falciparum* malaria to be responsible for 515 (range 300-660) million cases in 87 malaria endemic countries. It was also estimated that 1 million people die in Africa from infection by *P. falciparum* alone. The incidences of *P. falciparum* were found to be 50% higher globally than what was estimated by the World Health Organization (WHO; Snow *et al.* 2005).

Population groups most at risk to malaria are pregnant women, who lose their acquired immunity and children (Barnes *et al.*, 2008). Malaria is the biggest killer of children of any infectious disease in the world at present and unfortunately the mortality from malaria appears to be increasing (Olliaro, 2005). Malaria poses a great socio-economic burden on endemic countries and it is estimated to cost these already poverty stricken countries nearly US\$ 12 billion annually (Cambell *et al.*, 2004). The lack of a vaccine and the rapid emergence and spread of drug resistant strains of *P. falciparum* therefore necessitate the development of new drugs and the identification and validation of new parasite-specific therapeutic targets (Hyde, 2005).

The current study focused on the polyamine pathway, which when altered, disturbs

normal cell proliferation and differentiation. Intervention of the polyamine biosynthetic pathway has widely been studied in parasitic protozoa. The only success story up to date is the inhibition of ODC by DFMO in *T. brucei,* which is used in the effective treatment of African sleeping sickness (Muller *et al.*, 2001). This gives hope that altering the polyamine pathway in the malaria causative agent *P. falciparum,* would yield the same result. *P. falciparum* has been shown to be sensitive to the inhibition of ODC at several stages of the life-cycle with moderate to promising results (Muller *et al.*, 2008).

The essential nature of SpdSyn in parasitic protozoa is reflected in the importance of its product, spermidine. Spermidine contributes to the general role of polyamines, however, most importantly spermidine is a precursor for the modification and activation of eIF-5A and in trypanosomes for the biosynthesis of the gluthathione mimic, trypanothione (Byers *et al.*, 1992; Muller *et al.*, 2003; Kaiser *et al.*, 2003). Some effects of polyamine biosynthesis inhibitors have therefore been attributed to the accumulation of unmodified eIF-5A due to spermidine depletion, for example, in the suppression of multi-drug resistant HIV-1 replication (Bachrach and Abu-Elheiga, 1990; Park *et al.*, 1993; Schafer *et al.*, 2006). Null mutants of the SpdSyn of protozoan *L. donovani* have also shown that this enzyme is absolutely essential for the survival of lower eukaryotes (Guo *et al.*, 1999; Roberts *et al.*, 2001; Jin *et al.*, 2002). Recently SpdSyn has been validated genetically to be a drug target against *T. brucei* (Taylor *et al.*, 2008). The inhibition of PfSpdSyn by a known inhibitor, cyclohexylamine, has been shown to block *P. falciparum* growth by depleting the endogenous polyamine pools (Kaiser *et al.*, 2001). Parasites inhibited by 4MCHA could not be significantly rescued by the addition of exogenous spermidine (Haider *et al.*, 2005). This justified the subsequent investigation of PfSpdSyn to identify novel lead compounds.

The first objective of this study was to further exploratory work performed in the construction of a homology model for PfSpdSyn using TmSpdSyn and AtSpdSyn as templates. No realistic binding poses of the natural occurring substrates could be obtained during docking studies. Subsequently, the substrate analogue AdoDATO was converted

in silico into putrescine and dcAdoMet and subjected to a MD simulation to determine their respective binding modes. From the MD simulation a water molecule was observed within the catalytic center and was suspected to play a role in the mechanism of action of PfSpdSyn. This observation was further investigated in the current PhD study. More MD simulations were performed and a novel mechanism of action was derived. This was supported by observations from SpdSyn crystal structures of various organisms. During the preparation of the manuscript (Burger *et al.* 2007; Chapter 2) the first crystal structure of PfSpdSyn was released (Aug 2006). A structural superimposition revealed an excellent correlation between the Cα-backbone of the crystal structure and model of the PfSpdSyn (RMSD 0.594Å). Later, the release of a crystal structure of the human SpdSyn resolved in complex with putrescine and MTA, revealed that no water molecule is present within the catalytic center upon binding of putrescine as was suggested by MD simulations. This conflicting result could be explained by the conformational changes that the active site of PfSpdSyn undergoes in the absence of ligands and was later confirmed by the release of the apo-structure. It was therefore concluded in this study that differing results obtained by comparison of *in silico* and experimental studies were not due to the quality of the homology model, but due to the nature of the approach followed to determine the binding modes of putrescine and dcAdoMet.

The second and main objective of this study was the identification of novel hit compounds against PfSpdSyn. The discovery of drugs/hits against new drug targets can be derived by following either a random or rational approach. In the past the discovery of drugs has often depended on serendipity and has included strategies such as *in vitro* HTS. The HTS technology is limited to big pharmaceutical companies due to the high cost involved in screening of targets. HTS is also limited by its high attrition rate, with a hit-rate of between 0.01-1% for compounds screened (Neamati and Barchi, 2002). The advent of the "omics" era has seen a shift to the more rational approach known as SBDD (Sawyer 2006). Recent advances in the biological field have resulted in vast resources of information that need to be mined on order to better understand biological systems for

application in the drug design process (Sawyer, 2006).

Although the random and rational approaches differ significantly, the techniques used to accomplish the goal of finding new drugs or lead compounds can frequently be used interchangeably. Therefore the array of both different computational and experimental disciplines available to the drug discovery process should be seen as a multifaceted discipline which can contribute to the early stages of this process, regardless of the approach followed. As a result smart drug discovery platforms have been put in place in most pharmaceutical companies (Sawyer, 2006).

A smart discovery platform for lead discovery was developed for this study and involved the development of a dynamic receptor-based pharmacophore model as well as a knowledge-based rational design strategy. These two strategies were selected to complement each other in the selection of compounds to be tested *in vitro*. The use of the pharmacophore models to identify lead compounds has become increasingly popular over the last decade and has been shown to be a reliable method in the drug discovery process. The methodology followed in this study allows for the incorporation of protein flexibility within the drug design process which has been a long standing challenge. The development of a dynamic receptor-based pharmacophore model resulted in a wealth of information of the chemical space of the active site and should be used in a knowledge-based rational design strategy.

The development of a dynamic receptor-based pharmacophore was subdivided into four stages: 1) structure quality assessment, 2) phase space sampling, 3) negative image construction and 4) hit analysis. From the structure quality assessment stage it was concluded that both the PfSpdSyn structures, PDBid 2I7C and PDBid 2PT9, used in this study are of high quality and that confidence can be taken in conclusions drawn from them. The dimer of PfSpdSyn (PDBid 2I7C) was subjected to a 5ns MD simulation to sample the conformational space of PfSpdSyn using NPT conditions during the phase space sampling stage. The MD trajectory was clustered to identify the most

representative conformations and captured within it the dynamic behavior of the protein. Six structures from the MD simulation were selected as well as the three monomers of PfSpdSyn (PDBid 2PT9) from which the DPMs could be derived. These nine structures are representative of the phase space of PfSpdSyn. Negative image construction was done for the nine structures by performing MIF analysis, a grid-based method, to identify binding hotspots and subsequently to derive PhFs. The active site of PfSpdSyn was subdivided into four binding regions (DPM1-DPM4) to allow for the identification of fragments binding within these specific binding regions. DPMs were constructed from the identified PhFs representative of the chemical space of each binding region. In the hit analysis stage the DPMs were screened against a drug-like subset of the ZINC database to identify compounds containing the pharmacophore of interest. Compounds were filtered based on best fit values and docked to PfSpdSyn from which the selection of compounds to be screened *in vitro* was made.

The knowledge-based rational design strategy involves a literature review of known inhibitors, visual inspection of protein ligand-complexes, determination of structure activity relations, common feature identification of known ligands, the incorporation of knowledge of the binding site derived from MIF analysis, the *in silico* construction of compounds and the subsequent docking thereof. Compounds showing both good docking energies and binding poses are selected for similarity searches of commercially available compounds.

Both the development of a dynamic receptor-based pharmacophore model and the knowledge-based rational design strategy were followed in this study. A total of nine compounds were selected to be tested *in vitro*. These compounds were specifically selected to complement each other and to test for specific characteristics of the active site. Two compounds were shown to cause a significant reduction in the specific activity of PfSpdSyn *in vitro*. Compound ZINC02582060 was identified using a DPM representative of the DPM1 binding region and showed a 16% reduction in PfSpdSyn activity at 100μM. The compound NAC showed a reduction of 86% of the PfSpdSyn activity at 100μM. NAC

was designed to bind in both the dcAdoMet and putrescine binding cavities by bridging the highly negatively charged catalytic center. Furthermore, NAC was designed using the knowledge-based rational design strategy by incorporating knowledge gained from MIF analysis. Kinetic analysis was performed for NAC and it was found to have a K_i of 2.8 μM. The binding in both the dcAdoMet and putrescine cavities was confirmed by a Lineweaver-Burk plot which suggested that NAC does not only competes with putrescine for binding but also with dcAdoMet. The aminopropyl chain of NAC is proposed to form hydrogen bonds with residues His103, Asp127 and Asp196, which is similar to the aminopropyl chain of dcAdoMet. The cyclohexylamine nitrogen forms hydrogen bonds with Ser197 and Tyr102. Ser197 is part of the gate-keeping loop and it is proposed that this interaction keeps the loop closed over the active site for longer resulting in stronger inhibition. Two modifications have been suggested for the lead optimization of NAC. These compounds include NACD and NACDSW and it is proposed that the interactions these compounds make with the gate-keeping loop will increase its binding affinity. NACDSW also exploits the chemical space normally occupied by the positively charged sulphur of dcAdoMet. It can also be concluded from this study that the putrescine binding cavity can accommodate a piperidine ring when its ring nitrogen is oriented away from the catalytic center.

During the course of this study more crystal structures of PfSpdSyn were released. It is proposed that knowledge from these structures should be incorporated in the receptor-based pharmacophore model developed for PfSpdSyn in further studies. The new pharmacophore model should be screened against chemical databases and knowledge from the current study should be incorporated during compound selection for *in vitro* screening.

The development of a dynamic receptor-based pharmacophore model together with a knowledge-based rational design strategy has been shown to be an effective approach in the identification of novel hit compounds in the presence of a 3D target structure. The identification of NAC, a novel lead for PfSpdSyn should first be tested *in vitro* to determine its IC$_{50}$. The discovery of NAC has provided a new scaffold with room for

many chemical modifications to optimize its interaction and increase its binding affinity with PfSpdSyn. NAC has the potential to be used in treatment of malaria and should be further exploited by developing a set of lead compounds. After having derived a lead series it should be optimized using both *in vivo* and *in silico* methods to obtain an optimized lead with the desired properties to enter pre-clinical trials. The discovery of this compound is not limited to use in PfSpdSyn but may potentially play a important role in treatment of African trypanosomiases. Therefore, it can be concluded that a new lead compound NAC (K_i 2.8 μM) has been identified for PfSpdSyn with great potential for lead optimization.

The development of a dynamic receptor-based pharmacophore model for PfSpdSyn has seen the establishment of this methodology in the Bioinformatics and Computational Biology Unit, Department of Biochemistry at the University of Pretoria. This paves the way for more studies on both malaria and other drug targets using SBDD. It should be emphasized that the power of *in silico* techniques such as the development of a receptor-based pharmacophore model lies in the collaboration with research groups capable of performing complementary *in vivo* and *in vitro* studies.

Therefore, it can be concluded that this study has not only contributed to the fight against malaria but has also contributed in building the national capacity for SBDD in South Africa. It is thus envisaged that this study will further both the fight against malaria and science as a whole in South Africa.

Summary

Malaria affects the daily lives of more than 2 billion people worldwide and has been estimated to result in 300-500 million clinical cases annually leading to approximately 2 million deaths, mainly caused by the most virulent malaria species, *Plasmodium falciparum.* The lack of a vaccine and the rapid emergence and spread of drug resistant strains of *P. falciparum,* necessitate the development of new antimalarials and the identification and validation of new parasite-specific therapeutic targets.

Numerous studies directed at interfering with the polyamine biosynthetic pathway in *P. falciparum* have shown its potential as a target for the development of a new class of antimalarials. The essential nature of *P. falciparum* spermidine synthase (PfSpdSyn), an enzyme in the polyamine pathway of the parasite warranted the further investigation to find novel lead compounds. The high cost and attrition rate of drug discovery has resulted in the implementation of smart drug discovery platforms in both academia and industry. The strategy implemented in this study involved the development of a dynamic receptor-based pharmacophore model (DPM) of PfSpdSyn complemented by a knowledge-based rational design strategy.

The use of pharmacophore models to identify lead compounds has become increasingly popular over the last decade and has been shown to be a reliable method in the drug discovery process. The development of a DPM allows for the incorporation of protein flexibility within the drug design process. This methodology results in a wealth of information of the chemical space of the active site and was incorporated in designing new inhibitors against PfSpdSyn using a knowledge-based rational design strategy. The active site of PfSpdSyn was subdivided into four binding regions (DPM1-DPM4) to allow

for the identification of fragments binding within these specific binding regions. DPMs representative of the chemical characteristics of each binding region were constructed and subsequently screened against the drug-like subset of the ZINC database. From the screens a total of nine compounds were selected for *in vitro* testing, complementing each other in exploring specific active site binding characteristics. From these compounds a new lead compound N-(3-aminopropyl)-cyclohexylamine (NAC; K_i 2.8 μM) was identified for PfSpdSyn. NAC was specifically designed to bind in both the putrescine and decarboxylated adenosylmethionine cavities by chemically bridging the catalytic center and was confirmed by kinetic studies. NAC shows great potential for lead optimization to increase its binding affinity. This study then paves the way for lead optimization and possibly the development of a novel antimalarial.

The development of a DPM for PfSpdSyn has seen the establishment of this methodology in the Bioinformatics and Computational Biology Unit, Department of Biochemistry at the University of Pretoria. It can be concluded that the development of a DPM complemented by a knowledge-based rational design strategy is an effective approach for the identification of novel lead compounds in the presence of a 3D target structure. This paves the way for more studies on both malaria and other drug targets using DPMs.

Bibliography

Agomo, P. U., Meremikwu, M. M., Watila, I. M., Omalu, I. J., Odey, F. A., Oguche, S., Ezeiru, V. I. and Aina, O. O. (2008) Efficacy, safety and tolerability of artesunate-mefloquine in the treatment of uncomplicated *Plasmodium falciparum* malaria in four geographic zones of Nigeria. *Malar J* **7**, 172.

Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Aide, P., Sigauque, B., Milman, J., Mandomando, I., Bassat, Q., Guinovart, C., Espasa, M., Corachan, S., Lievens, M., Navia, M. M., Dubois, M.-C., Menendez, C., Dubovsky, F., Cohen, J., Thompson, R. and Ballou, W. R. (2005) Duration of protection with RTS,S/AS02A malaria vaccine in prevention of *Plasmodium falciparum* disease in Mozambican children: single-blind extended follow-up of a randomised controlled trial. *Lancet* **366**, 2012–2018.

Antosiewicz, J., McCammon, J. A. and Gilson, M. K. (1996) The determinants of pKas in proteins. *Biochemistry* **35**, 7819–7833.

Aponte, J. J., Aide, P., Renom, M., Mandomando, I., Bassat, Q., Sacarlal, J., Manaca, M. N., Lafuente, S., Barbosa, A., Leach, A., Lievens, M., Vekemans, J., Sigauque, B., Dubois, M.-C., Demoitie, M.-A., Sillman, M., Savarese, B., McNeil, J. G., Macete, E., Ballou, W. R., Cohen, J. and Alonso, P. L. (2007) Safety of the RTS,S/AS02D candidate malaria vaccine in infants living in a highly endemic area of Mozambique: a double blind randomised controlled phase I/IIb trial. *Lancet* **370**, 1543–1551.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. and Yeh, L.-S. L. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115–D119.

Assaraf, Y. G., Golenser, J., Spira, D. T., Messer, G. and Bachrach, U. (1987) Cyto-