

Chapter 2

***In silico* modelling of PfPdx1 to facilitate selection of potential R5P-active site inhibitors**

2.1 Introduction

2.1.1 Drug discovery

There are several different screening methods for the identification of novel antimalarial compounds and these include whole cell or phenotypic screening, target-based screening and virtual-based screening. Phenotypic screening or *in vitro* whole cell assays have been more successful in identification of chemical leads since 1) usually no additional modifications are required to improve pharmacokinetics or permeability of identified compounds and 2) an unbiased approach is employed [99, 146]. Another advantage of whole cell screens, which could lend support to its success, is the fact that the complete parasite biology is represented and therefore interrogated during screens. Antimalarials such as CQ and other 4-aminoquinolines affect haemozoin polymerisation, a cellular process, rather than a specific protein. One of the disadvantages of this approach is the fact that the molecular target is not always known. Resistance surveillance therefore becomes more intensive and tailoring the compound for improved efficacy is harder. Fragment-based libraries compliment phenotypic screens and uses small molecules (100 – 250 Da) at relatively high concentrations to provide an initial skeleton for additional chemical optimisations [147]. This drug design methodology argues that as the complexity of a ligand increases, the chance match to a potential drug target decreases [148]. A library of complex molecules is therefore not as likely to succeed in having the desired effect on a drug target compared to smaller and simpler molecule types.

Target-based screening involves the pre-selection and validation of essential parasites proteins, which are then either subjected to HTS or structure-based screening methods (if the target structure is available). In the laboratory target-based screening requires the establishment of sensitive *in vitro* assay procedures and is often hampered by poor quality and low quantities of recombinantly expressed proteins from malaria parasites. Moreover, compounds identified in such screens may not always have the required potency *in vivo*, necessitating chemical optimisation to either improve efficacy and/or pharmacokinetic and pharmacodynamic properties [99]. Pursuing compounds identified from target-based inhibition screens are often not successful as these fail in whole cell activity assays, even after successive compound optimisations [99, 146, 149]. It has been proposed that chemical optimisation to improve cell permeability and pharmacokinetics results in fewer leads

compared to whole-cell phenotypic screening approaches [99, 149]. Newer drug-delivery methods may, however, complement such shortcomings [99, 149]. One advantage of target-based drug identification is that these approaches can more effectively benefit from ligand-target binding information, such as NMR or crystallographic evidence of inhibitor interactions, resulting in rapidly optimised leads.

The advancement of computational techniques has dramatically aided in drug discovery and virtual-based screening approaches have become more prominent largely due to lower costs and being less time consuming compared to both whole cell and target-based screens [150, 151]. Virtual-based screening can either be used to extract useful information from ligands or the actual protein target structure or receptor. Ligand-based methods can be used to identify the 3D shape, chemical similarity or pharmacophore features of known active compounds and aid in the selection of analogous compounds [150]. [150]. When the target structure is known, receptor or structure-based computational or virtual screening methods can be used [150]. Structure-based drug design in turn aims to predict complementary interactions ligands have with a receptor. This enables the rationalization of binding modes and comparisons of the structure activity relationships (SAR) ligands possess. Structure-based methods include molecular docking simulations, which aim to identify low-energy binding partners, as well as pharmacophore approaches, which describe the necessary chemical features compounds should possess to bind effectively. An example of a successful structure-based study by Haque *et al.* identified potent inhibitors with K_i -values in the nanomolar range against two aspartyl protease plasmepsin proteins of the malaria parasite [152]. The computational technique was first utilised to derive building blocks for the library synthesis from a protein homologue cathepsin D, after which these compounds were tested on plasmepsin [152, 153]. Another study focussed on identifying more specific inhibitors against cysteine protease homologues from *Leishmania donovani*, *Trypanosoma cruzi* and *P. falciparum* (cysteine proteases termed falcipain-2 and falcipain-3) [154]. Computational libraries of around 335 000 compounds were filtered and docked into the *T. cruzi* structure as well as both falcipain-2 and falcipain-3 homology models. Several inhibitors with low micromolar efficacy were identified against falcipain-2 and the *L. donovani* proteins, and an additional pharmacophore representative of these proteins was created and deemed useful for additional interrogations of the active site [154].

A pharmacophore model is an ensemble of chemical features required by a compound, protein or cellular component for molecular recognition [155]. The pharmacophore includes chemical moieties on a molecule that are needed for interactions with a protein receptor and *vice versa* the amino acid residues of a protein that enable recognition of a molecule or cellular binding partner. Ligand-based pharmacophores can be created by overlaying known substrates, inhibitors or

products and extracting the shared chemical characteristics [155]. Structure-based pharmacophore screening approaches aim to map compounds to the chemical environment of the target active site. Structure-based pharmacophore mapping entails identifying complementary chemical features and their spatial arrangement in the active site. The receptor-ligand complex is often used to locate the active site, alternatively atoms in cavity regions and interaction maps are generated to probe potential ligand binding site. These pharmacophores are then screened using virtual screening and searching approaches. This approach helps to discover structurally diverse compounds which are capable of targeting the receptor protein and aims to provide a comprehensive repertoire of molecules which are effective. Structure-based pharmacophores are generated by identifying interactions between substrates or products and their respective protein receptors (Figure 2.1). A limitation to structure-based drug design could be the lack of crystallographic data of structures, and also validated targets, and this is often overcome by using homology modelling.

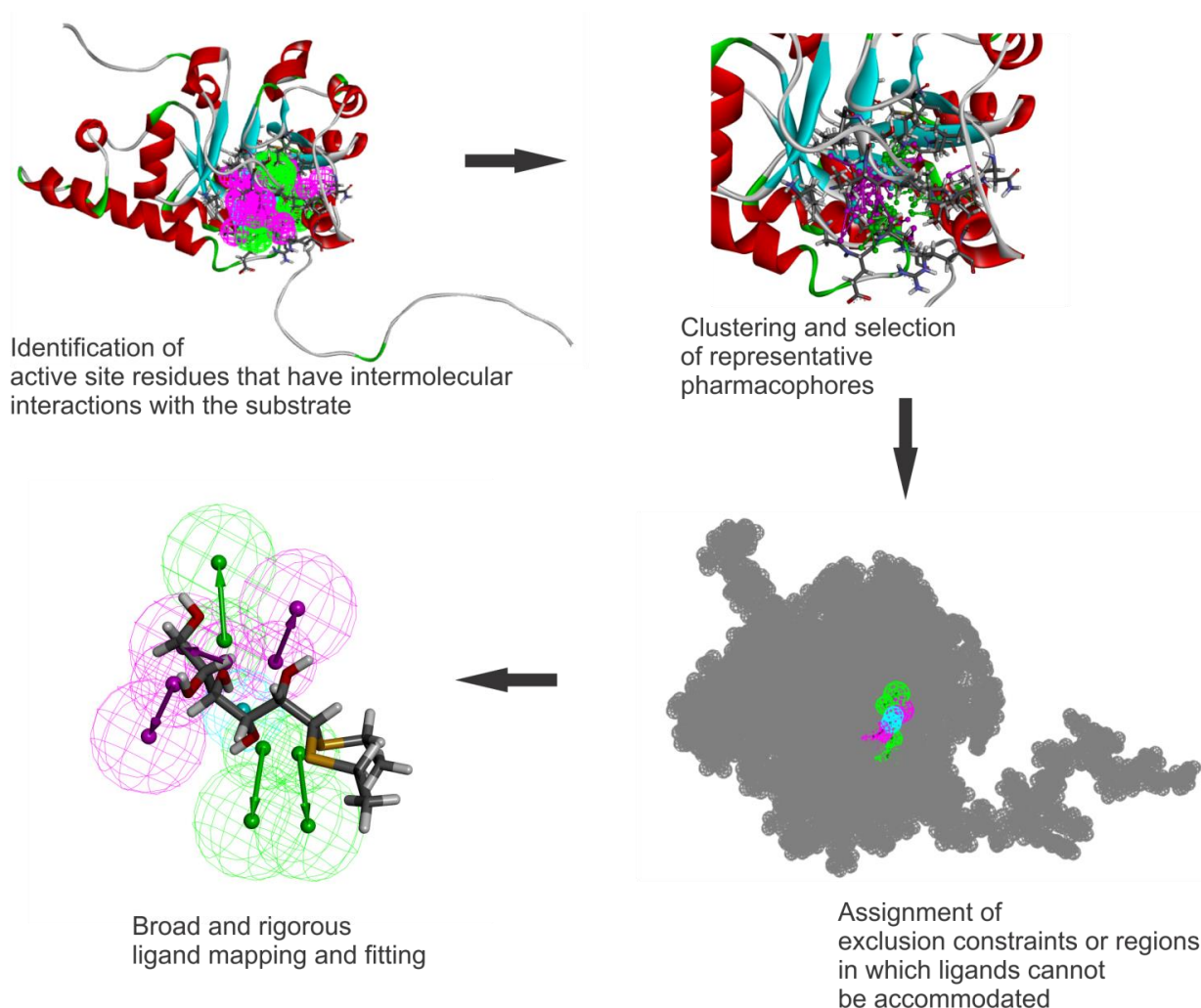


Figure 2.1: Workflow for the identification of ligands from a structure-based pharmacophore. Protein structures or homology models are used to extract pharmacophores from the active site. Disallowed regions are assigned in the hypothetical pharmacophore model based on the location of other amino acid residues that are not directly associated with the substrate. Ligands are then finally mapped to the pharmacophores and aligned with complementary chemical moieties.

2.1.2 The Pdx1 component of PLP synthase

The PLP synthase is composed of two proteins; Pdx1 and Pdx2 (Figure 2.2). In *B. subtilis*, *G. stearothermophilus* and *Thermotoga maritima* it was shown that the PLP synthase complex forms a large dodecameric ring structure, composed of two hexameric rings, each hexameric ring consisting of six Pdx1 subunits surrounded by six Pdx2 proteins [119, 156, 157]. Assembly of the *P. falciparum* PLP synthase is initiated when inactive *Pf*Pdx2 monomers decorate the two *Pf*Pdx1 hexameric rings to form a dodecamer of 424 kDa [133]. Structural characterisation of the plasmodial PLP synthase has highlighted important aspects of assembly and exploitable characteristics for the development of novel therapeutics [133]. During assembly of PLP synthases Pdx1 activates the Pdx2 glutaminase activity, resulting in the hydrolysis of L-glutamine, thereby releasing ammonia that is channelled to Pdx1 subunits.

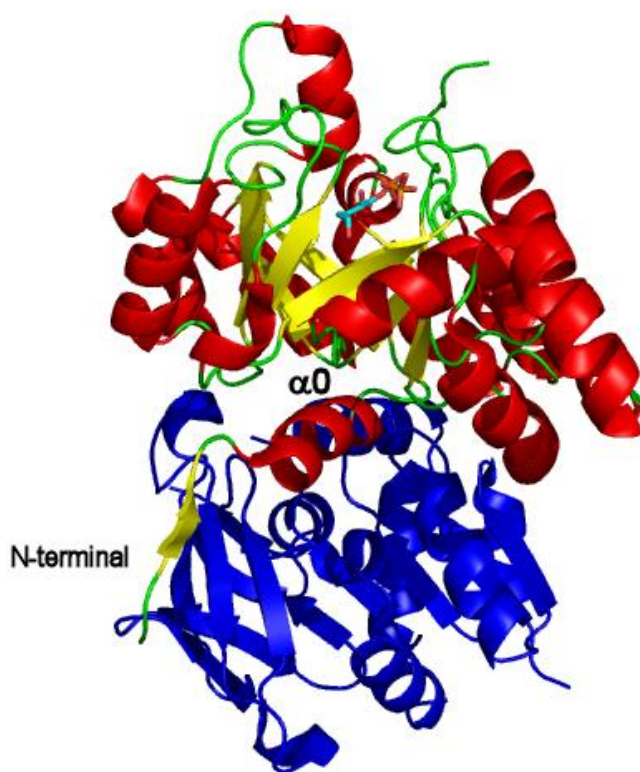


Figure 2.2: The heterodimeric structure of *T. maritima* PLP synthase (PDB: 2ISS). *Tm*Pdx1, coloured in red and yellow, associates with *Tm*Pdx2, coloured in blue [157]. *Tm*Pdx1 has TIM (α/β)₈ barrel fold architecture with eight central parallel β -sheets (yellow) alternated with by eight α -helices (red), with some modifications. The N-terminal β -sheet and the α -helical region (α_0 or α_N) of *Tm*Pdx1 contacts the *Tm*Pdx2 monomer and have been shown to stimulate glutaminase activity in *Tm*Pdx2 [156, 157].

Pdx1 proteins are highly conserved, however the N-terminal region, spanning the first 20 amino acids, differs markedly between species and plays an important role in PLP production [158]. The Pdx1 N-terminal stimulates Pdx2 activity in the assembled heterodimeric complex, and this region in *Pf*Pdx1 is believed to contribute half of the total interactions between heterodimers [159]. The

crystal structure of *TmPdx1* in complex with *TmPdx2* revealed ordered interactions promoted by the Pdx1 N-terminal region, including residues 8 - 16 which form an α -helix, termed α_0 or α_N (Figure 2.2) [157]. The *B. subtilis* Pdx1 (*BsPdx1*) N-terminal region, composed of an additional short β -strand and α -helix, is essential for *BsPdx2* binding and interaction [156]. Both *TmPdx1* and *BsPdx1* N-terminal regions have predominately polar interaction with Pdx2, involving both acidic and basic residues [157, 159]. (Figure 2.2) [157]. The *B. subtilis* Pdx1 (*BsPdx1*) N-terminal region, composed of an additional short β -strand and α -helix, is essential for *BsPdx2* binding and interaction [156]. Both *TmPdx1* and *BsPdx1* N-terminal regions have predominately polar interaction with Pdx2, involving both acidic and basic residues [157, 159].

The *PfPdx1* enzyme performs the bulk of the catalytic formation of PLP. The enzyme combines R5P, G3P and ammonia derived from active Pdx2 monomers, to synthesise PLP. The reaction mechanism was already comprehensively studied and elucidated at the initiation of this study [160-162]. These studies eluded to the complex imine chemistry involved in different active sites, however at the time only one site with the R5P substrate was definitively known from crystallographic evidence of *TmPdx1*, including evidence from mutational studies [157]. The R5P binding site is also termed the P1 site. Pdx1 enzymes can also bind D-ribulose-5-phosphate (Ru5P) in the P1 site, and this pentose sugar is also incorporated into the final PLP molecule, which is discussed in more detail in Chapter 3. Two additional regions in Pdx1 were later identified which were shown to accommodate PLP and G3P, and were termed the P2 and P3 sites respectively [156, 163, 164].

Contributions of amino acid side-chains in the *TmPdx1* R5P-binding or P1 site are shown in Figure 2.3. The first part of the Pdx1 reaction mechanism is believed to involve imine formation of R5P with the internal K residue [160]. The involvement of a K82 residue was confirmed in *TmPdx1*, in which it was observed to be covalently bound, through the formation of an imine, to the C2 carbon of Ru5P (Figure 2.3) [157]. In the recently solved *Plasmodium berghei* structure of Pdx1 (*PbPdx1*) it was shown that unlike in the *TmPdx1* structure, the K84 residue was covalently linked to the C1 carbon of R5P [164]. The *PbPdx1* and *TmPdx1* structures are therefore progressive snapshots of the initial R5P-binding stages, respectively. Other residues that are important in the R5P-binding site include a D25 in *TmPdx1* (D27 in *PbPdx1*) involved in H-bond contacts with the hydroxyl groups located on the C2 or C3 carbon of R5P and Ru5P, respectively [157, 164]. Residues G215, G236 and S237 in *TmPdx1* (Figure 2.3) are H-bonded to the phosphate group of Ru5P, and are believed to facilitate anchoring of the substrate [157]. Residues in *PbPdx1* ascribed with similar roles include G217, S239 and G238 and additionally, G156 [164].

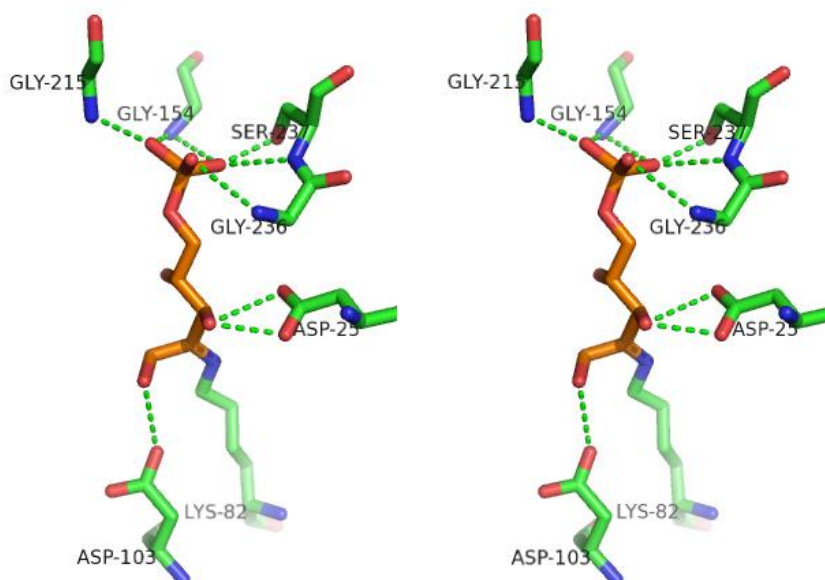


Figure 2.3: Stereoview of the *TmPdx1* R5P-active site. The figure shows residues involved in hydrogen bonding to the Ru5P molecule, which include K82 forming a covalent imine bond with the substrate, as well as a series of glycine residues involved in stabilization of the phosphate moiety [157].

From a structure-based pharmacophore drug design approach, *PfPdx1* is an attractive drug target as this protein is highly conserved and crystallographic information of Pdx1 homologues are available. The high degree of sequence identity shared amongst Pdx1 proteins enables quality homology modelling, and allows more truthful predictions regarding ligand-receptor binding interactions [165]. *PfPdx1* homology models were generated and five representative structure-based pharmacophores were extracted from the R5P-binding site. These pharmacophores were used for the virtual screening of 2 million ligands to identify potential inhibitory compounds. Results from the virtual screening were confirmed with molecular docking to support pharmacophore-identified compounds.

2.2 Methods

2.2.1 Homology modelling

Alignment of templates and molecular modelling was performed using the Discovery Studio (DS) suite (v3.0, Accelrys Software Inc.) unless specified differently. Multiple sequence alignments (MSA) to compare the different templates were generated using the BLOSUM 62 pairwise alignment scoring matrix. The gap open penalty was set to 11, with an extension penalty of 1. *Pf*Pdx1 homology models were generated using *T. maritima* (*Tm*Pdx1, PDB accession: 2ISS) and *Thermus thermophilus* (*Tt*Pdx1, PDB accession: 2ZBT) Pdx1 as templates. At the time this study was initiated the *P. berghei* structure of Pdx1 (*Pb*Pdx1, PDB accession: 4ADU) was not available.

Models were generated using aligned templates in MODELER within the Build homology models protocol [166]. MODELER optimises spatial restraints to derive models from provided template structures. Two separate homology models were generated, one in which the Ru5P substrate from the 2ISS template was copied, the other model was built containing the (4S)-2-methyl-2,4-pentanediol (MPD) from the 2ZBT template. These structures were labelled *Pf*Pdx1_Ru5P and *Pf*Pdx1_MPD, respectively. The 2ZBT template crystal structure was solved with several water molecules in near vicinity of the MPD ligand. These were labelled as H₂O51, H₂O58, H₂O95, H₂O102, H₂O124 and H₂O139 in the 2ZBT structure, and were additionally included in the *Pf*Pdx1_MPD homology model, which was labelled *Pf*Pdx1_MPD_W. Five versions of the models were generated and the model with the lowest probability density function (PDF) total energy was used in further comparisons. Both the PDF total energy and PDF physical energy was calculated during this process. Additionally the discrete optimized protein energy (DOPE) score was computed for each homology model. Structural comparisons and superimpositions were made using both the DS (Accelrys Inc.) and the Java-based RCSB PDB comparison tool [167] with the Java implemented CE algorithm (jCE, version 1.1) [168]. Ramachandran plots were generated using PDBsum [169, 170] and the online resource PROCHECK was used to stereochemically assess the qualities and geometry of the homology models [171].

2.2.2 Protein minimisation

The *Pf*Pdx1 homology models were prepared for pharmacophore mapping by adding hydrogens using the CHARMM forcefield, after which the models were subjected to energy minimisation steps [172]. The CHARMM forcefield tool is a flexible molecular dynamics and modelling program which assigns atomic charges to atoms and residue types to residues in the models in preparation for

energy minimisation. In energy minimisations with the CHARMM minimisation algorithm 1000 steps of Steepest Descent were used with an RMS gradient tolerance of 3 kcal/mol.Å and a distance cutoff value of 14 Å for nonbonded interaction pairs. The maximum steps during minimisation were limited to 200 and the minimisation energy change was set to 0. This minimisation was followed by a conjugate gradient minimisation with the same parameters except for the RMS gradient which was set to 0.1 kcal/mol.Å. The substrate within the minimised structure was used to assign a binding site sphere with a radius of 8Å.

2.2.3 Interaction generation

The minimised homology models were subjected to the interaction generation protocol in the DS. This approach probes the receptor and identifies the interaction map or pharmacophore features in the assigned binding sphere based on amino acid properties. The position of the individual ligands in the models allowed the creation of an active site sphere. The sphere radius was, in each case, set to 8.0 Å. The water molecules in the *PfPdx1_MPD_W* model were included as receptor components. Water molecules form hydrogen bond (H-bond) networks which facilitate ligand-receptor interactions [173]. Although dynamic, inclusion of these waters provided additional features in the ensemble of pharmacophores not derived from active site residues. In the *PfPdx1_MPD_W* model these water were H₂O51 (O,H1,H2), H₂O58 (O3,H3,H4), H₂O95 (O1,H5,H6), H₂O102(O5,H7,H8), H₂O124(O2,H9,H10) and H₂O139 (O4,H11,H12). Ligands were deleted from the models before initiation of the protocol. Running parameters were altered with the density of lipophilic sites set to 5 and the density of polar sites was also set to 5.

2.2.4 Pharmacophore clustering

From interaction generation between 20 and 70 different pharmacophore features were identified from the respective homology models. These pharmacophore features were hierarchically clustered into a distance matrix allowing a pharmacophore cluster centre to be selected based upon the distance from another feature. A custom feature pharmacophore according to interactions displayed for the Ru5P ligand in the homology model, was also created. Exclusion pharmacophore features were assigned to residues located outside a 5Å radius from the identified pharmacophore features.

2.2.5 Database searching and ligand pharmacophore mapping (LPM)

The pharmacophores were screened against the drug-like subset of the Zinc database comparing 2 million ligands [174]. The database was generated as a multi-conformer composite database using catDB with a maximum of 250 representative conformations (Accelrys Software Inc.). In the database searching algorithm the best searching method was used which adds additional flexibility to the ligands during fitting. The drug-like database was screened against 5 pharmacophore models and the combined hits produced from the 5 pharmacophores were further submitted for scoring based-pharmacophore mapping. All of the 5100 ligands identified from database searching, were re-submitted for LPM in which all the conformations of the ligands were used. A rigid fitting method was employed with the energy threshold of the ligands set to 20 kcal/mol. During mapping the atoms of the ligands are optimally fitted to the pharmacophore features using an algorithm established by Kabsch *et al.* [175]. The fit value is computed using both the numerical weight assigned to individual pharmacophore features and the distance of the ligand atoms which are fitted to the centre of a corresponding pharmacophore feature. The individual pharmacophore features of *PfPdx1* pharmacophores were equally weighted, with each feature assigned a numerical value of 1. For stringent selection of ligands at least $n - 1$ pharmacophore features had to match in order to map the ligands (where n represents the total number of pharmacophore features in the representative model).

Chemical diversity was determined using Chemmine tools (<http://chemmine.ucr.edu>), a freely available online tool used to clusters compounds based on pairwise similarities which are defined using Atom Pair and Tanimoto descriptors. Atom Pair descriptors refer to the atom organization in terms of the paths around non-hydrogen atoms in a molecule, whereas Tanimoto descriptors refer to the proportion of features shared amongst compounds, as retrieved from <http://chemmine.ucr.edu/iframe/work/intro/tutorial>. The 50-best ligands identified from LPM were clustered based on their single-linkage pairwise similarities.

2.2.6 Docking

To reinforce the selection of ligands against *PfPdx1*, 300 of the best-hit compounds, based on their pharmacophore fit-values, were further subjected to molecular docking simulations. Molecular docking simulations using LigandFit were conducted using a dreiding energy forcefield, [176, 177]. Molecular docking simulations using LigandFit were conducted using a dreiding energy forcefield, [176, 177]. During conformational searching of the R5P-active site 1500 Monte Carlo trials were

used to iteratively generate ligand conformations, with the dielectric constant set to 1.0, and a non-bonded cut-off distance of 10.0 Å. Docking was performed with a RMS threshold of 2.0 Å for ligand/binding site matching. Ligand poses were ranked according to DockScore functions based on forcefield approximations, taking into account the van der Waals and electrostatic interaction energy between ligand and receptor, as well as the ligand internal energy. Ligand poses were additionally evaluated using the LigScore empirical scoring functions [178]. This defines van der Waals interaction between ligands and the receptor, derived using polar attraction, and desolvation penalties incurred when ligands bind to a protein receptor. R5P was used as control in cross-docking to validate ligand docking and scoring. Following the release of the *PbPdx1* structure in 2012 [164], additional docking was performed on the *PfPdx1* homologue. The *PbPdx1* structure was prepared by removing the covalent bond between K84 and R5P, followed by minimisation as described above. Post-minimisation, R5P was selected to assign a binding site and sphere radius, similar to the docking simulations described above. K84 was maintained in a protonated state during the docking simulations (predicted pK_a of 12.1).

2.3 Results

2.3.1 Alignments and template selection

A multiple protein sequence alignment (MSA) of the template Pdx1 crystal structures and the *Pf*Pdx1 homologue showed a 60% and 61% sequence identity with *T. maritima* Pdx1 (*Tm*Pdx1, PDB accession 2ISS) and *T. thermophilus* Pdx1 (*Tt*Pdx1, PDB accession 2ZBT), respectively (Figure 2.4). The sequence similarity was 79% and 77% compared to *Tm*Pdx1 and *Tt*Pdx1, respectively. After the release of the *P. berghei* Pdx1 crystal structure (*Pb*Pdx1, PDB accession 4ADU) an alignment was performed and showed a 86% sequence identity with *Pf*Pdx1, however at the time of homology modelling this structure was not available [164].

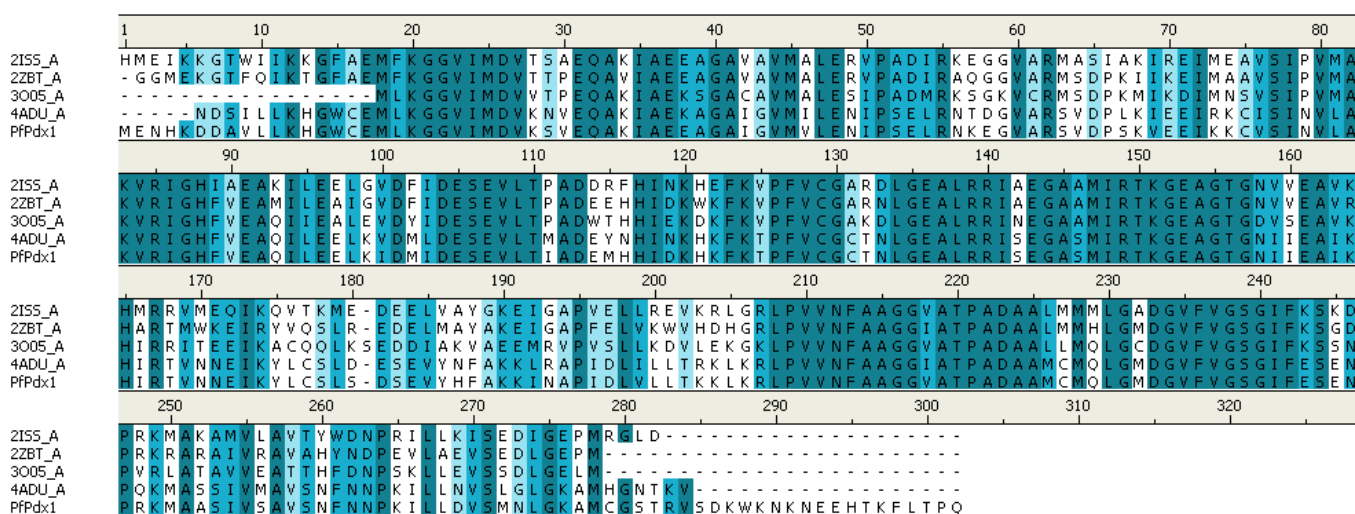


Figure 2.4: Multiple protein sequence alignment (MSA) generated for Pdx1 homologues. Aligned crystal structures of *T. maritima* (2ISS), *T. thermophilus* (2ZBT), *P. berghei* (4ADU), *S. cerevisiae* Pdx1 (3O05) and *Pf*Pdx1 (pyridoxine biosynthetic enzyme pdx1 homologue, NCBI reference sequence XP_966196.1). Conserved residues indicated in dark blue, with residues with similar properties indicated in light blue.

The secondary structure topology profile of *Tm*Pdx1 and *Tt*Pdx1 are shown in Figure 2.5, together with the calculated residue conservation. These topology profiles were shown to be similar with some differences noted at the N-terminal region. *Tt*Pdx1 has more extensive β -sheet formation followed by a shorter α -helical region compared to *Tm*Pdx1 (Figure 2.5 A). Residues which form part of the R5P-binding site are indicated in the topology profiles (Figure 2.5). These included D25, M44, P50, K82, G154, G214 and G215 (as labelled in *Tm*Pdx1).

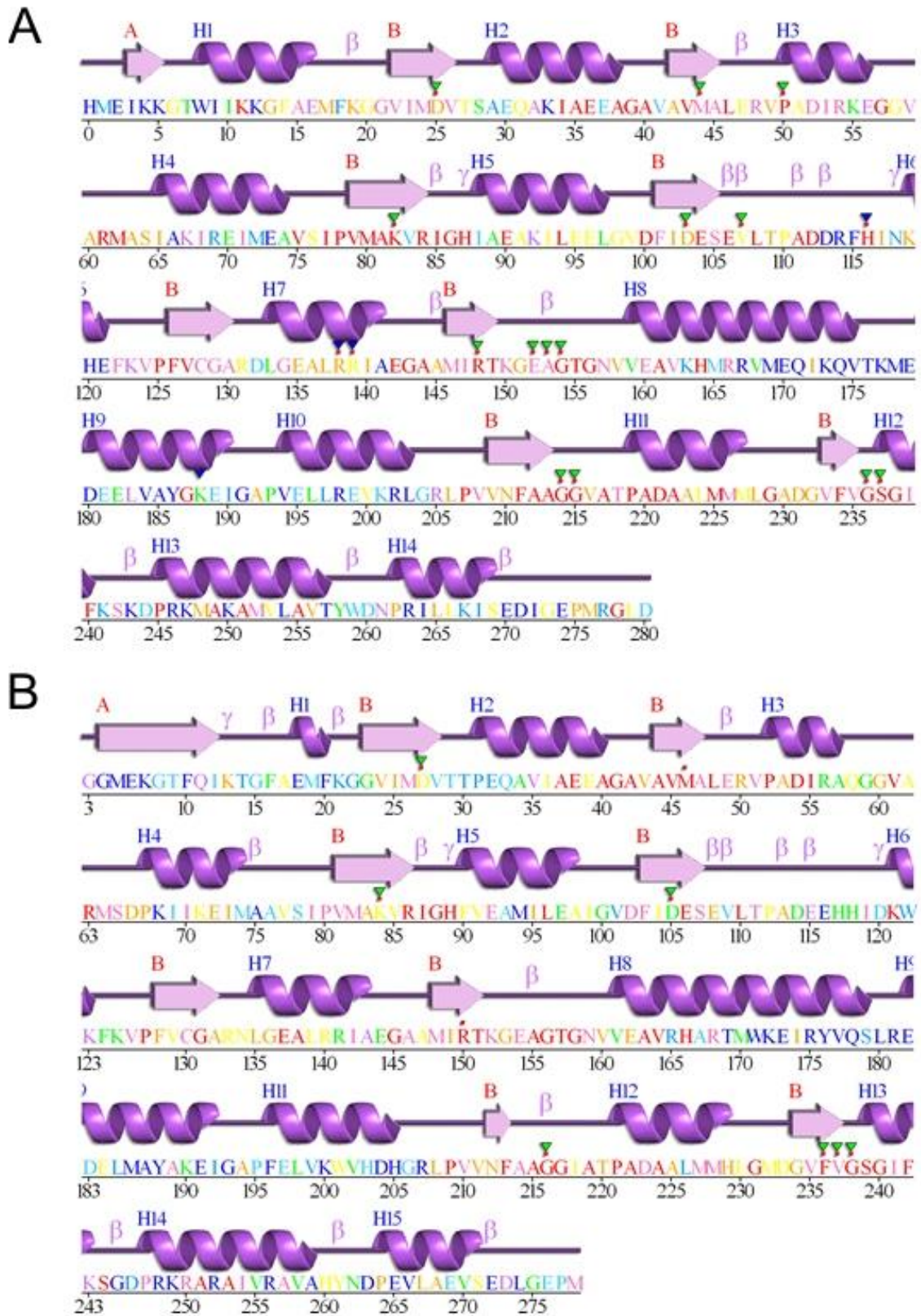


Figure 2.5: Secondary structure topology diagrams of A) *TmPdx1* and B) *TtPdx1*. Topology diagrams were generated in PDBsum [169, 170]. Secondary structures such as α -helices are labeled with H, whereas β -sheet are labeled with B. β - and γ -turns are indicated by β and γ respectively. Active site residues with predicted interactions with ligands are indicated by red dots above the residue. Additional green triangles indicate that the substrate is part of the defined active site record in the PDB file. Residue conservation calculated using ConSurf-DB server is indicated by colour, in which least conserved residues are blue, somewhat conserved residues are light blue to green, and most conserved residues are shades of red.

A structural comparison between A chains of the templates, *TmPdx1* and *TtPdx1*, revealed 74% sequence identity and 86% sequence with the C_α root mean square deviation (RMSD) of 1.0 Å (Figure 2.6 A). Residues surrounding the R5P substrate (5 Å radius) had a main chain RMSD (N, C_α , C, O) of 0.89 Å (Figure 2.6 B). Structural comparison of K82, D25, M44, R148 and G215 showed a good correlation with these residues adopting similar positioning (Figure 2.6 B). The mechanistically implicated K82 residue in *TmPdx1*, covalently bonded to the Ru5P ligand, was shown to have a similar orientation to K83 in *TtPdx1* (Figure 2.6 B). The residue forms a H-bond with MPD in *TtPdx1*. Residue D25 of *TmPdx1* was also shown to H-bond with the hydroxyl group of Ru5P, compared to the equivalent D27 residue in *TtPdx1* which had two predicted H-bonds toward the hydroxyl groups of MPD. The S237 (S239 in *TtPdx1*) and P50 (P52 in *TtPdx1*) residues between the two templates differed significantly and is located more than 3.4 Å from each other (Figure 2.6 B).

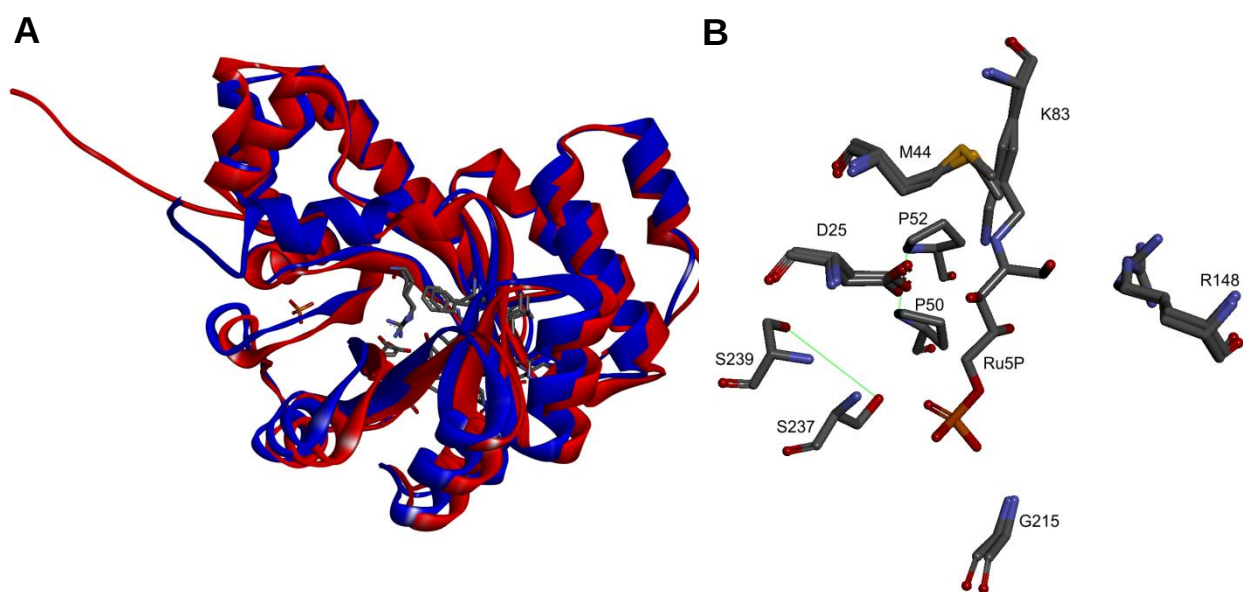


Figure 2.6: Comparison of *TtPdx1* (PDB: 2ZBT) to *TmPdx1* (PDB: 2ISS). **A)** Superimposition of the proteins revealed a C_α RMSD of 1.0 Å. **B)** Superimposition of R5P-binding site residues revealed conservation of residue locations, with a main chain RMSD of 0.89 Å. Markedly different residues were P50 and S237 which were located more than 3.4 Å from each other (*TmPdx1* labelling).

2.3.2 Homology modelling of *PfPdx1*

Four homology models were generated based on the template sequence alignments. These models were generated each to include different substrate and water molecules from the different template structures. Three of the models contained the MPD ligand from the *TtPdx1* structure

(*PfPdx1_MPD*, *PfPdx1_MPDsco* and *PfPdx1_MPDw*) whereas the other contained the Ru5P ligand from *TmPdx1* (termed *PfPdx1_Ru5P*). The models were subjected to protein quality assessment in which the PDF total energy, physical PDF and discrete optimized protein energy (DOPE) scores were calculated before being used in further studies (Table 2.1). The PDF total energy was determined and gives the model-derived pseudo energy terms as well as the stereochemical pseudo-energy terms. In addition the physical PDF energy was determined and describes the summed energies of stereochemical pseudo-energy terms, consisting of valence angles and torsion angles, valence bonds, improper torsion angles and soft sphere repulsions.

Table 2.1: Properties of homology models built using MODELER in DS.

	PDF total energy ^a	PDF physical energy ^a	DOPE score ^b
<i>PfPdx1_MPD</i>	7355.68	761.53	-31510.11
<i>PfPdx1_Ru5P</i>	6918.45	727.15	-31585.13
<i>PfPdx1_MPDw</i>	7355.68	761.53	-31510.11
<i>PfPdx1_MPDsco</i>	7355.68	761.53	-31510.11

^a PDF total and physical energies are unit-less descriptors of the restraint violations in the models, in which a higher comparative value indicates greater violations

^b DOPE score or relative stability of the model is used to compare different states of the same model, whereas the absolute unit-less score is not meaningful due to the algorithms not normalizing for protein size [179].

In addition, the individual PDF scores for individual residues of models reflecting their steric pseudo-energy environments were calculated and are shown in Figure 2.7. Residues such as P51 and S238 in the *PfPdx1_MPD* models had high PDF total scores indicating steric collisions which were further resolved using the CHARMM energy based minimisation procedure to resolve any steric clashes between side chain orientations. Resulting models such as *PfPdx1_MPDsco* had similar residue PDF total energy line plots as for *PfPdx1_Ru5P* (results not shown).

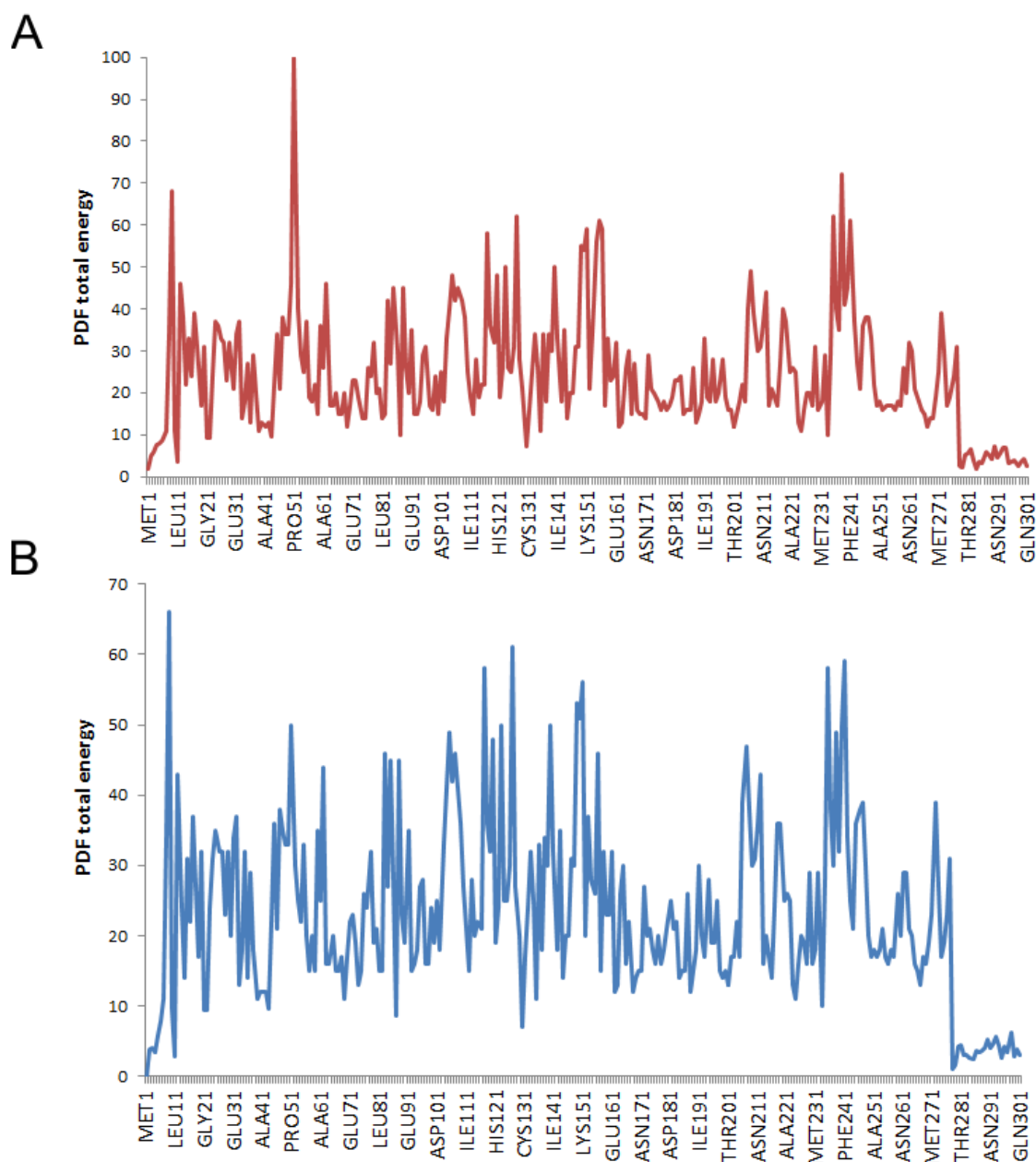


Figure 2.7: Line plot of the PDF total score in homology model (A) *PfPdx1_MPD* and (B) *PfPdx1_Ru5P*. Residues that had high PDF total scores include P51 and S238. The former P51 residue was calculated with high bond angle potential.

Moreover, Ramachandran plots were used to verify residue phi- (ϕ) and psi- (ψ) angle discrepancies. Ramachandran plots were generated for *PfPdx1_MPD* and *PfPdx1_Ru5P* (Figure 2.8), whereas the remaining models *PfPdx1_MPDsco* and *PfPdx1_MPDw* had similar plots and are not shown. The residue ϕ and ψ -angles were clustered into distinct regions in the plots, corresponding to secondary structural elements these occurred in [180]. There were 21 glycine residues and 9 proline residues in the *PfPdx1* models. Glycine has greater backbone conformational flexibility due to the lack of side-chain, therefore allowed in additional regions of the plots, whereas

proline is rotationally constrained and limited to certain ψ -angles [180, 181]. No residues were found in disallowed regions with only two residues present in the generously allowed regions; E58 and K99 (Figure 2.8). The E58 and K99 residues were located on solvent exposed loop regions in *PfPdx1*, and were distant from the active site, and would not have influenced interpretation of the R5P-binding site pharmacophore.

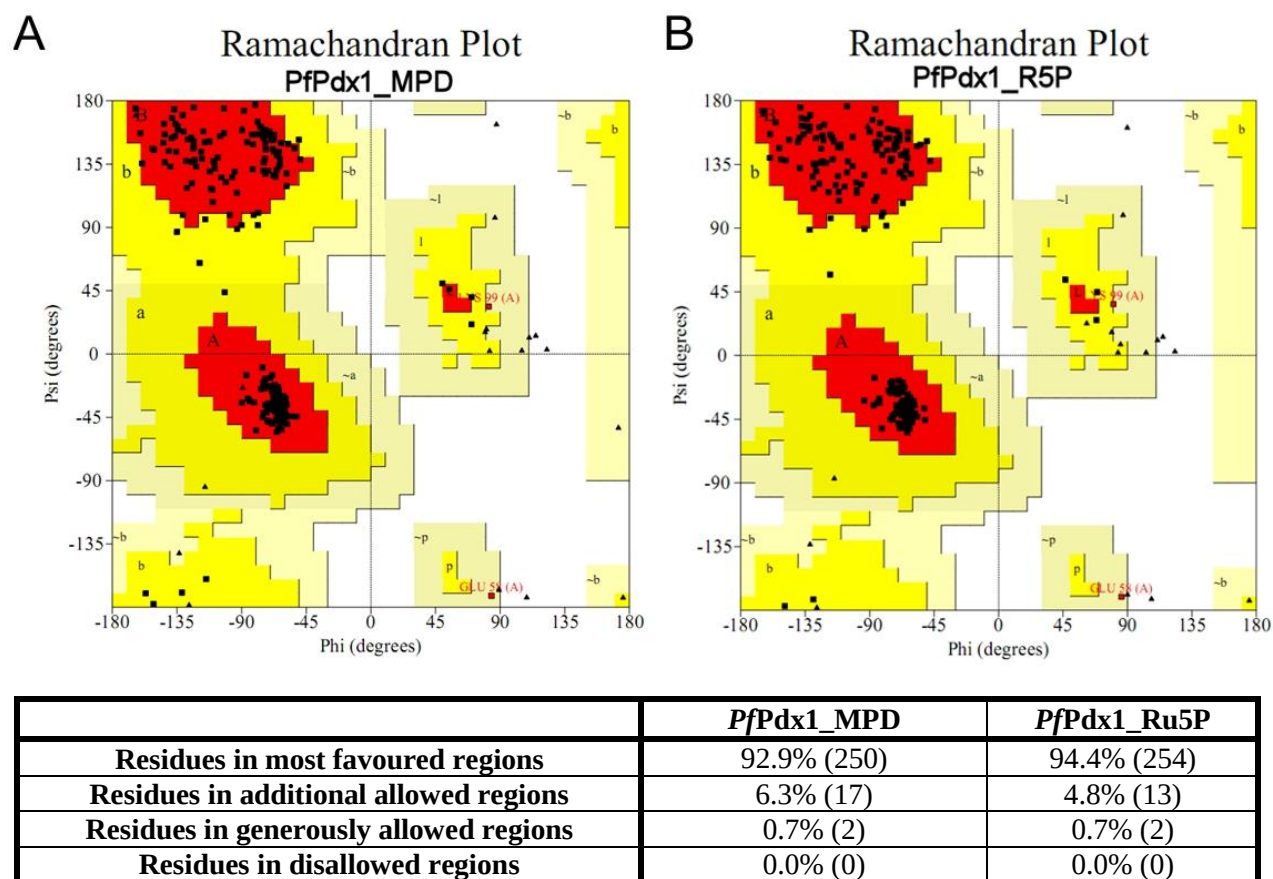


Figure 2.8: Ramachandran plots of *PfPdx1*_MPD (A) and *PfPdx1*_Ru5P (B). The plots represent the ϕ and ψ angle combinations of the residues in the *PfPdx1* models. Areas coloured in red represent most favoured angle distribution regions, areas coloured in yellow and light yellow represent additionally and generously allowed angle distribution regions, respectively. Individual amino acid residues are indicated by black squares, or red squares when within generously allowed regions. Glycine residues are indicated as triangles. Values in the table represent the percentage of residues in respective regions with the number of residues in parenthesis.

The C_{α} RMSD's of *PfPdx1* models prior to minimisation compared to the template structures revealed values of between 0.9 and 0.6 Å (Table 2.2). The comparative C_{α} RMSD value between our models and the later released *PbPdx1* was 0.6 Å (Table 2.2).

Table 2.2: RMSD's of *Pf*Pdx1 homology models. The homology models were compared using the RCSB PDB comparison tool, using an established jCE algorithm [167]. The *Pf*Pdx1 models were compared to template and *P. berghei* Pdx1 structures.

	C _α RMSD compared to <i>Tm</i> Pdx1 (2ISS) (Å)	C _α RMSD compared to <i>Tt</i> Pdx1 (2ZBT) (Å)	C _α RMSD compared to <i>Pb</i> Pdx1 (4ADU) (Å)
<i>Pf</i> Pdx1_MPD	0.38	0.87	0.68
<i>Pf</i> Pdx1_Ru5P	0.27	0.92	0.64
<i>Pf</i> Pdx1_MPDw	0.38	0.87	0.68
<i>Pf</i> Pdx1_MPDsco	0.38	0.87	0.68

In addition the C_α and main chain atoms (N, C_α, C, O) RMSD values for the active site (within 5 Å radius of Ru5P/MPD) were calculated and compared to determine C_α backbone and side chain atom orientations (Table 2.3). The RMSD's of the *Pf*Pdx1 models revealed smaller deviations compared to *Tm*Pdx1 than to *Tt*Pdx1. The *Pf*Pdx1 active site residues were comparable to the *Pb*Pdx1 active site, with RMSD between 0.5 – 0.6 Å, indicating that good quality models were generated.

Table 2.3: RMSD's of residues surrounding the R5P-binding site ligand in *Pf*Pdx1 homology models. Un-minimised *Pf*Pdx1 models were compared to Pdx1 templates and *P. berghei* Pdx1 using residues surrounding the ligand in the active site. The C_α RMSD is reported and main chain atom RMSD is given in parenthesis.

	RMSD compared to <i>Tm</i> Pdx1 (2ISS) (Å)	RMSD compared to <i>Tt</i> Pdx1 (2ZBT) (Å)	RMSD compared to <i>Pb</i> Pdx1 (4ADU) (Å)
<i>Pf</i> Pdx1_MPD	0.37 (0.36)	0.70 (0.73)	0.62 (0.66)
<i>Pf</i> Pdx1_Ru5P	0.22 (0.24)	0.76 (0.76)	0.59 (0.64)
<i>Pf</i> Pdx1_MPDw	0.37 (0.36)	0.70 (0.73)	0.62 (0.66)
<i>Pf</i> Pdx1_MPDsco	0.37 (0.36)	0.70 (0.73)	0.62 (0.66)

R5P-binding site residues in the generated *Pf*Pdx1 homology models were spatially comparable to the *Tt*Pdx1 and *Tm*Pdx1 template structures (Table 2.3). In Figure 2.A1 of Appendix 2 residues aligned better with the *Tm*Pdx1 templates compared to *Tt*Pdx1 supporting the RMSD calculations in Table 2.3. The P52 residue of *Pf*Pdx1, which differed markedly between the two template structures, adopted a conformation similar to that of P51 in the *Tm*Pdx1 structure (Figure A2.2, Appendix 2). These results suggested that the high degree of sequence conservation between *Pf*Pdx1 and templates led to the creation of structurally similar homology models.

The overall topology of the *Pf*Pdx1 structure had a (α/β)₈ barrel fold architecture with eight central parallel β-sheets alternated with by eight α-helices (Figure 2.8). This is in agreement with topologies of *Tm*Pdx1, *Bs*Pdx1 and *G. stearothermophilus* Pdx1 [119, 156, 157]. Additions to the (β/α)₈ barrel structure include the N-terminal α-helix, αN, as well as α-helices α2a, α6a, α6b and

$\alpha 8a$ located in between β/α pairs (Figure 2.8 A) [157]. A right handed 3_{10} helix, with characteristic 3-residue H-bonding pattern, was located between $\beta 8$ and $\alpha 8$ (Figure 2.9 A).

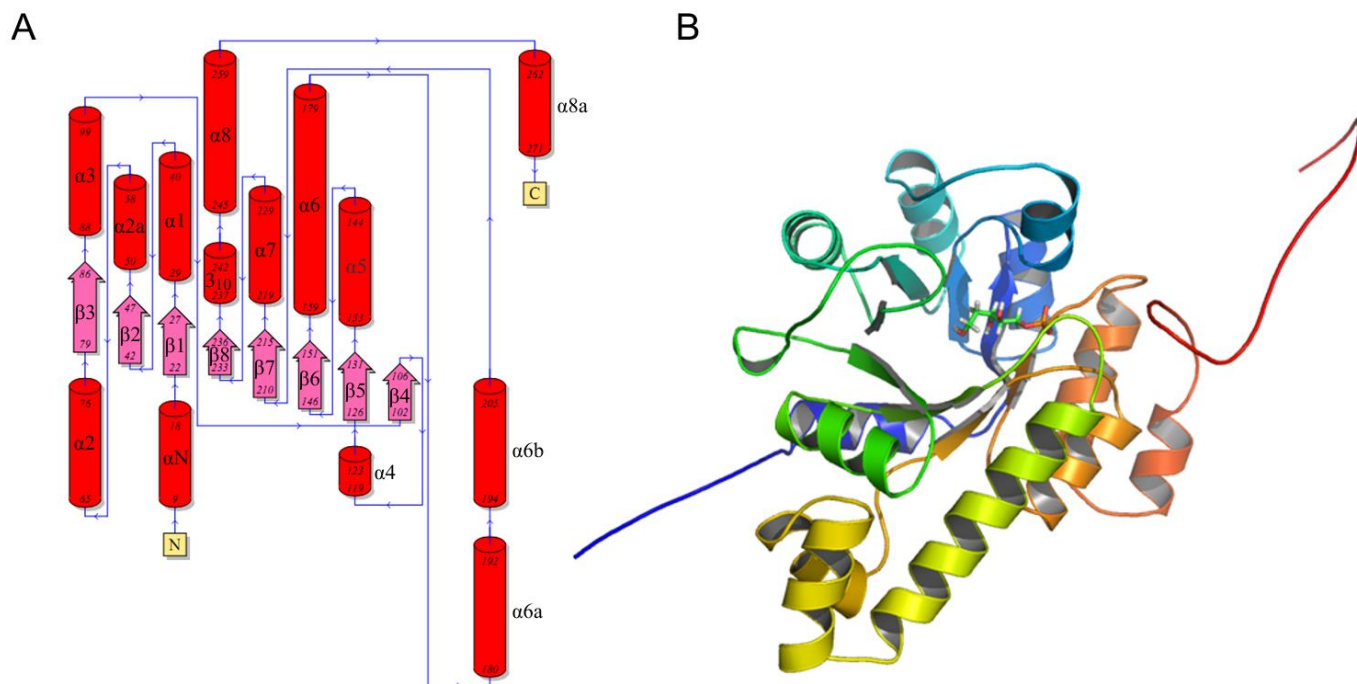


Figure 2.9: Topology and homology model of *PfPdx1_Ru5P*. **A)** The topology of *PfPdx1* was comparable to *TmPdx1* with eight central parallel β -sheets alternated with by eight α -helices. The topology diagram was generated using PDBsum [169, 170]. **B)** Homology model of *PfPdx1* (model *PfPdx1_Ru5P*). The view is from the top down the β -barrel. The model is indicated with the N-terminal in blue and the C-terminal in red. The Ru5P ligand is shown as a stick line structure located in the β -barrel region.

The *PbPdx1* was previously shown to complex with *PfPdx2* resulting in the formation of chimeric PLP synthase dodecamers, which was shown to have comparable catalytic activities to *P. falciparum* PLP synthase [164]. The high degree of identity shared between the proteins (86%) therefore enabled *PfPdx2* proteins to recognise similar structural elements in *PbPdx1* required for dimerisation. The bacterial *TmPdx1* has an additional β -strand on the N-terminal preceding αN which undergoes β completion with a central β -sheet in *TmPdx2* upon association [157]. *PbPdx1* was shown to lack this secondary structural element, and interactions between *PbPdx1* and *PfPdx2* predominated from αN region [164]. Superimposition of *PfPdx1* homology models with *PbPdx1* showed corresponding residues from αN in *PfPdx1* could be involved with interactions *PfPdx2* (Figure 2.10). The *PfPdx1* homology models lacked the N-terminal β -strand. Residues from the *PbPdx1* αN -region, identical to corresponding residues in *PfPdx1*, which form part of the interface with *PfPdx2* include M19 (in *PbPdx1*) [164]. This residue in *PfPdx1* (M18) is therefore similarly expected to be involved in contacts with *PfPdx2* (Figure 2.10). *PfPdx1* has 20 additional amino acids in the C-terminal region, which could not be aligned with any template structures, and this

region remained unstructured in the models (Figure 2.10). The *PbPdx1* structure similarly had an extended C-terminal region, which did not adopt any secondary structural arrangement (Figure 2.10).

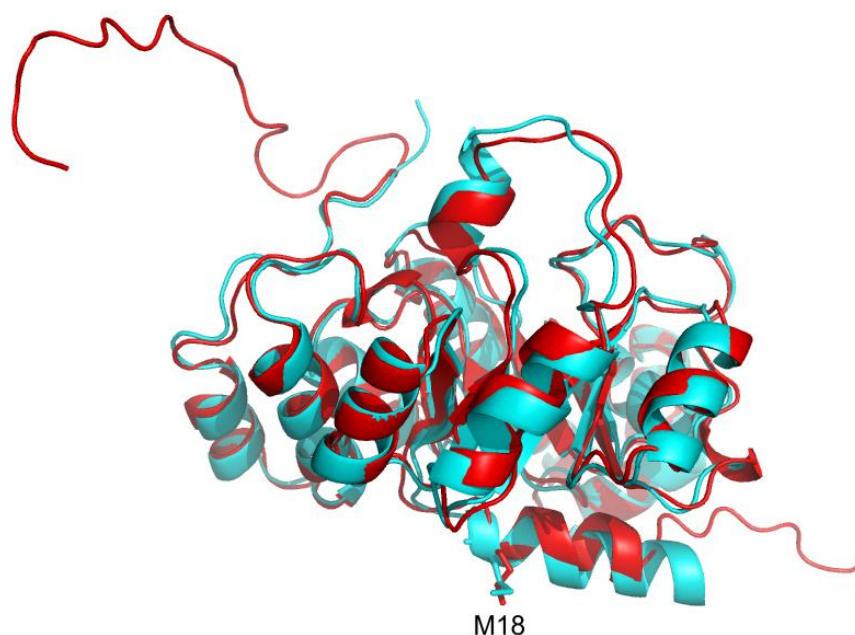


Figure 2.10: Overlay of *PfPdx1* with *PbPdx1*. Of the 13 residues in the N-terminal α -helix of *PfPdx1* (red), located in the lower part of the figure, were conserved and adopted similar orientations to *PbPdx1* (cyan). The conserved methionine residue (M18 in *PfPdx1*) proposed to be involved in contacts with the *PfPdx2* monomer in the assembled heterodimer, is indicated [164]. The unstructured C-terminal tails of *PfPdx1* and *PbPdx1* are shown on the left side of the figure.

2.3.3 Protein minimisation

Protein minimisation was performed to geometrically optimise the protein models, and remove steric clashes introduced by satisfaction of spatial restraints during superimposition model building by superimposition of ϕ and ψ angles of the template structures. The *PfPdx1*_MPDsco model was subjected to side chain optimisation to resolve steric clashes prior to protein minimisation, and had a lower initial potential energy compared to other models (Table 2.4). Comparing the final potential energies from the final structures with those of the initial structure indicated that some molecular geometric clashes and atomic overlaps were resolved during minimisation (Table 2.4). The final potential energy of the models was between -18800 and -19023 kcal/mol (Table 2.4).

Table 2.4: Properties of minimised *Pf*Pdx1 homology models.

	Initial potential energy (kcal/mol)	Final potential energy (kcal/mol)	Van der Waals energy (kcal/mol)	Electrostatic energy (kcal/mol)	Initial RMS gradient (kcal/(mol x Angstrom))	Final RMS gradient (kcal/(mol x Angstrom))
<i>Pf</i> Pdx1_MPD	14682.3	-18824.6	-1941.5	-19347.3	1748.3	1.28
<i>Pf</i> Pdx1_Ru5P	67073.7	-19358.4	-1923.7	-19822.6	8540.9	1.19
<i>Pf</i> Pdx1_MPDw	148887.6	-19023.2	-1913.7	-19583.9	20231.9	1.23
<i>Pf</i> Pdx1_MPDsc _o	-10065.7	-18816.6	-1884.9	-19420.8	12.0	1.28

Protein secondary structures of the minimised *Pf*Pdx1 homology model (Figure 2.11) were comparable to *Tm*Pdx1 (Figure 2.4). Secondary structures of other homology models are not shown, as these were identical to that of the *Pf*Pdx1_MPD model. In the assembled heterodimer, the N-terminal region of Pdx1 proteins spanning the first 20 amino acids interacts with Pdx2 and plays an important role in stimulating Pdx2 activity. This region, consisting of residues 8 – 16 form the α N α -helix, becomes ordered (adopts secondary structural arrangement) in *Tm*Pdx1 only in the presence of *Tm*Pdx2 monomers, [157]. The first 4 amino acids of the N-terminal region amongst different *Plasmodium* Pdx1 homologs have been shown to be variable, with structural evidence showing that the only ordered region in the N-terminal region of the *P. berghei* Pdx1 crystal structure consists of an α -helix spanning from residues D8 to M19 [158, 164]. Residues L10 – M18 in minimised *Pf*Pdx1 homology models adopted an α -helical arrangement (Figure 2.11 and Figure 2.12) similar to *Pb*Pdx1 structures.

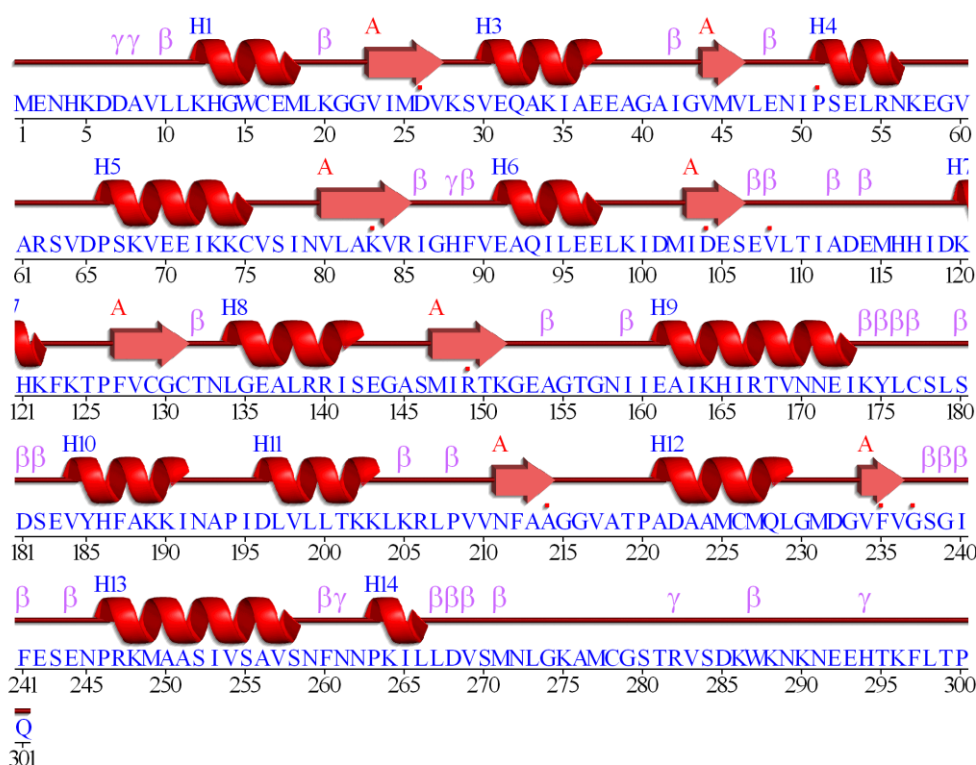


Figure 2.11: Secondary structure diagram of minimised *PfPdx1_MPD*. Predicted protein secondary structures were generated in PDBsum [169, 170]. Secondary structures such as α -helices are labeled with H, whereas β -sheet are labeled with A. β - and γ -turns are indicated by β and γ respectively.

The overall arrangement of secondary structural elements of the minimised *PfPdx1* model was similar to *TmPdx1* (Figure 2.12). Comparison to the *TmPdx1* showed similar secondary structural motifs were conserved. Some difference in the *PfPdx1_Ru5P* model included shortening of the central β_3 and β_4 sheets (Figure 2.12), which was not observed in the remaining *PfPdx1_MPD* models (data not shown). The β_3 regions of *PfPdx1* is composed of residues identical to that of *PbPdx1* which was shown to form a β -sheet from residue N80 to V85 [164]. These differences in secondary structure of this particular *PfPdx1* model were related to the algorithm used to predict secondary structure during model creation. RMSD comparison (see below) suggested that there was a good correlation between minimised *PfPdx1* models and the *TmPdx1* and *PbPdx1* structures.

Previous deletion mutagenesis of the C-terminal tail of *PfPdx1* showed that residues 270-279 were essential for the catalytic activity of *PfPdx1* and influenced R5P binding [182]. *PfPdx1* models were aligned with the chimeric *PbPdx1* structure in complex with *PfPdx2* (Appendix Fig A2.3). The C-terminal tail of *PfPdx1* models from residue M271 to T281 were shown to correlated with those from the *PbPdx1* structure. Residues in close proximity to the Ru5P molecule included M277, which was located 5.27 Å from the phosphate groups of the molecule.

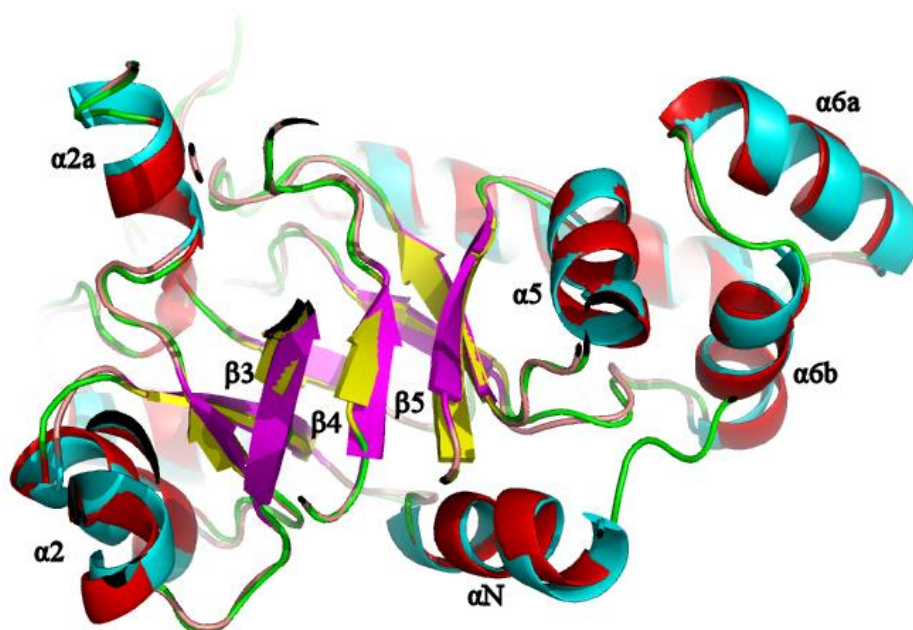


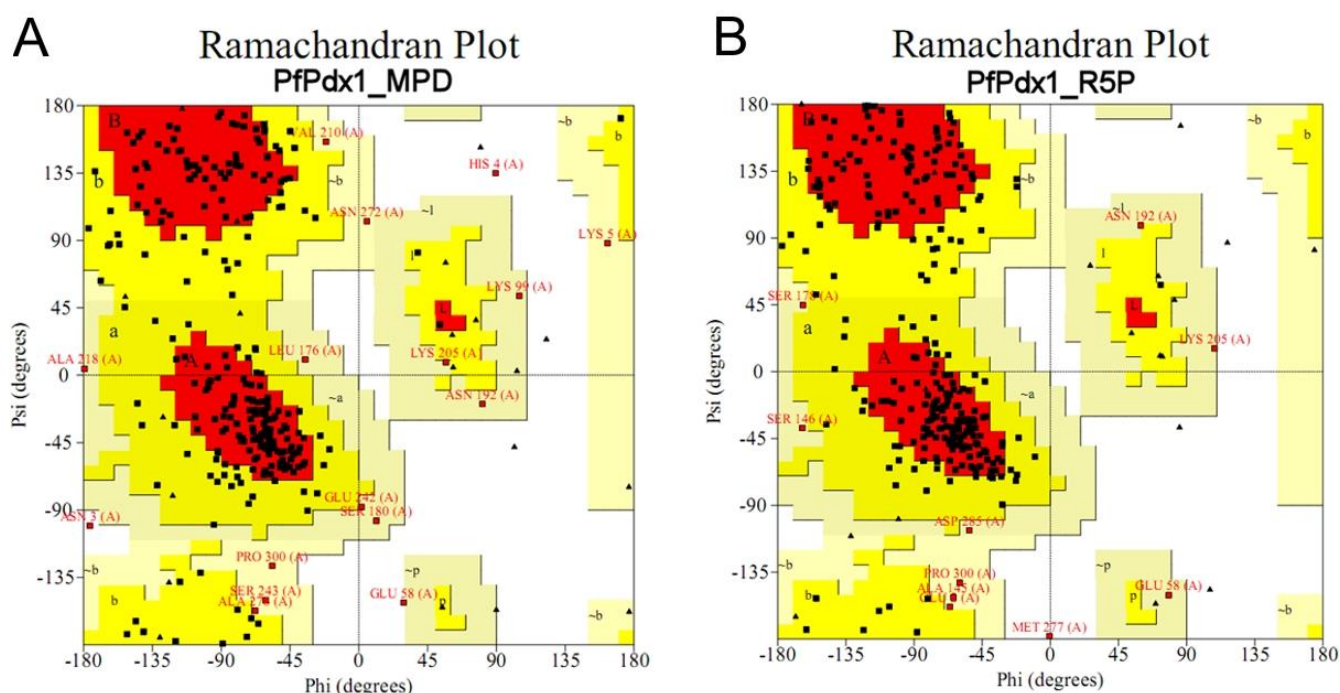
Figure 2.12: Side view of *PfPdx1_Ru5P* superimposed on the *TmPdx1* template structure. *PfPdx1_Ru5P* model (α -helices indicated in red, β -sheets indicated in yellow) was overlaid on the *TmPdx1* (α -helices indicated in cyan, β -sheets indicated in magenta). For ease of presentation the helix $\alpha3$ was omitted from the image.

RMSD comparisons of the minimised models to reported structures and templates showed a good correlation (Table 2.5). The deviations of minimised models were slightly increased from unminimised models, however were still below 1.0 Å for the *TmPdx1* template and *PbPdx1* structure.

Table 2.5: RMSD's of minimised *PfPdx1* models. The C_α RMSD's of *PfPdx1* models were compared to template and *P. berghei* Pdx1 structures using the RCSB PDB comparison tool [167] with the jCE algorithm.

	C_α RMSD compared to <i>TmPdx1</i> (2ISS) (Å)	C_α RMSD compared to <i>TtPdx1</i> (2ZBT) (Å)	C_α RMSD compared to <i>PbPdx1</i> (4ADU) (Å)
<i>PfPdx1_MPD</i>	0.79	1.09	0.96
<i>PfPdx1_Ru5P</i>	0.74	1.20	0.93
<i>PfPdx1_MPDw</i>	0.77	1.08	0.93
<i>PfPdx1_MPDsco</i>	0.80	1.14	0.98

Ramachandran plots after minimisation of *PfPdx1* homology models revealed a greater percentage of residues within generously allowed regions, as well as three residues having ϕ and ψ angles in disallowed regions (Figure 2.13). Ramachandran plots of additional homology models are given in the appendix (Appendix 2, Figure 2.A4).



	<i>PfPdx1_MPD</i>	<i>PfPdx1_Ru5P</i>	<i>PfPdx1_MPDsco</i>	<i>PfPdx1_MPDw</i>
Residues in most favoured regions	66.9% (180)	69.5% (187)	65.8% (177)	67.7% (182)
Residues in additionally allowed regions	27.5% (74)	27.1% (73)	28.6% (77)	28.3% (76)
Residues in generously allowed regions	4.5% (12)	3.0% (8)	4.5% (12)	3.0% (8)
Residues in disallowed regions	1.1% (3)	0.4% (1)	1.1% (3)	1.1% (3)

Figure 2.13: Ramachandran plots of *PfPdx1* homology models after protein minimisation. Values represent percentages with the number of residues in parenthesis. The plots of **A)** *PfPdx1_MPD* and **B)** *PfPdx1_Ru5P* models represent the ϕ and ψ angle combinations of the residues in the *PfPdx1* models. Areas coloured in red represent most favoured angle distribution regions, areas coloured in yellow and light yellow represent additionally and generously allowed angle distribution regions, respectively. Individual amino acid residues are indicated by black squares, or red squares when within generously allowed or disallowed regions. Glycine residues are indicated by triangles.

Some residues of *PfPdx1* models in generously allowed regions of the Ramachandran plots were part of disordered N- and C-terminal tails. These were not removed from models and therefore biased the Ramachandran plots. Of the residues in disallowed regions most occurred on the disordered N-terminal surface exposed region, such as N3 and H4, or the C-terminal tail, such as M277 (Figure 2.13, also Appendix Figure A2.3). The E58 residue, positioned on a surface exposed loop region between $\alpha 2a$ and $\alpha 2$, could not be minimised into alternative ϕ and ψ angles, and was also characterised into a disallowed region. Residues in surface exposed regions have higher propensities to exhibit conformations which result in characterization into disallowed

Ramachandran plot regions [183]. After energy minimisation, the orientation of residues in the active site was compared using RMSD calculations. The RMSD's of residues located in the R5P-binding active site of *PfPdx1* were below 1.0 Å compared to the *TmPdx1* and *PbPdx1*, suggesting conservation of the relative residue positioning (Table 2.6).

Table 2.6: RMSD's of the R5P-binding site of minimised *PfPdx1* models. The *PfPdx1* homology models were compared to templates and *PbPdx1* using residues surrounding the ligand in the active site (5 Å radius from ligand). C_α RMSD is reported and main chain atoms RMSD given in parenthesis.

	RMSD compared to <i>TmPdx1</i> (2ISS) (Å)	RMSD compared to <i>TtPdx1</i> (2ZBT) (Å)	RMSD compared to <i>PbPdx1</i> (4ADU) (Å)
<i>PfPdx1</i> _MPD	0.75 (0.83)	1.00 (1.08)	0.87 (0.96)
<i>PfPdx1</i> _Ru5P	0.70 (0.77)	1.06 (1.12)	0.87 (0.96)
<i>PfPdx1</i> _MPDw	0.76 (0.85)	1.00 (1.08)	0.86 (0.95)
<i>PfPdx1</i> _MPDsco	0.77 (0.87)	1.08 (1.14)	0.89 (0.99)

Residues in close proximity to the Ru5P ligand in *PfPdx1* superimposed and aligned well with equivalent residues in the *TmPdx1* structure (Figure 2.14 A). Spatial positioning of D26, K83, D104 and R149, located near the carbonyl component of the Ru5P ligand, were maintained (Figure 2.14 A). Residues involved in stabilization of the phosphate groups, such as G216, G237 and G155 (not shown) were comparable and aligned with each other (Figure 2.14 A). This implied that the R5P-binding site of the *PfPdx1* homology models were comparable to that of *TmPdx1*, and that similar residues are involved in stabilization of the substrate in *PfPdx1* [157]. The R148 residue from *TmPdx1* H-bonds with the C1 hydroxyl group of Ru5P [157]. The equivalent residue in *PfPdx1*, R149, had a similar spatial positioning, and could therefore be involved in H-bonds with the R5P terminal carbonyl group (Figure 2.14 A).

The P51 residue in *PfPdx1*, which was shown to differ in spatial location between the *TmPdx1* and *TtPdx1* template structures, had a similar positioning to that of *TmPdx1* (Figure 2.14 A). The equivalent P50 residue in *TmPdx1* had non-bonded contacts with the ligand and was predicted to contribute in hydrophobic stabilisation to Ru5P [157]. Similar roles could be assigned to the P51 residue in the *PfPdx1* model. Superimposition of *PfPdx1* with *PbPdx1* had comparable orientations of residues involved with contacts to the R5P substrate (Figure 2.14 B). The orientation of side chain atoms of the M45 residue in *PfPdx1* was different compared to both the *TmPdx1* template and *PbPdx1* structures, in which the sulphur atom was rotated 180° towards the ligand. This residue has been predicted to be involved in hydrophobic contacts towards the Ru5P ligand in *TmPdx1* [157]. The particular orientation the M45 residue in the *PfPdx1* models, with methyl groups in

approximately similar positions compared to *TmPdx1* and *PbPdx1* structures, is still expected to promote hydrophobic interactions with the Ru5P substrate.

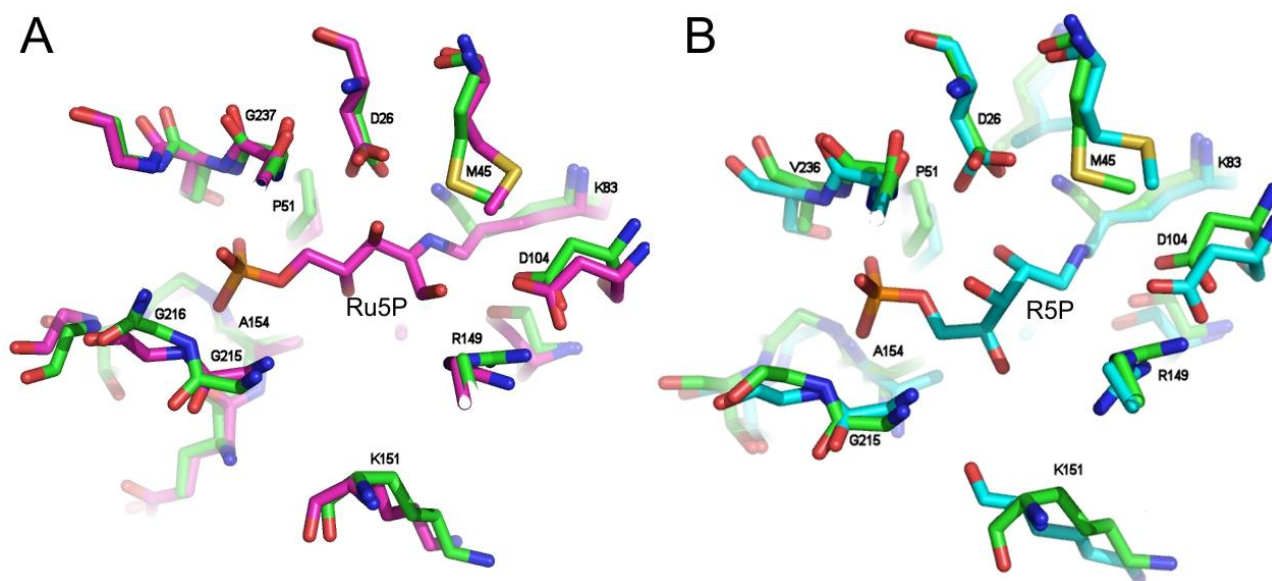


Figure 2.14: The R5P-binding site of *PfPdx1* homology models. **A)** The *PfPdx1*_Ru5P model (green) was compared to *TmPdx1* (magenta). The main chain RMSD of the active site residues was 0.77 Å suggesting good correlation of residue positioning. In contrast to *TmPdx1* the M45 residue of *PfPdx1* was orientated with the sulphur atom rotated towards the Ru5P substrate, with the methyl group adopting similar positioning to that of the *TmPdx1* structure. Residue labels are that of the *PfPdx1* model. The Ru5P substrate from *TmPdx1* is shown covalently bonded to K82 (*TmPdx1* numbering) in the centre of the figure. **B)** The *PfPdx1*_Ru5P model (green) superimposed with *PbPdx1* (cyan). Residues were labelled according to the *PfPdx1* structure. The R5P ligand from *PbPdx1* is shown in the centre of the figure.

2.3.4 Interaction generation and pharmacophore extraction

Pharmacophore chemical features were extracted from the R5P binding site of individual *PfPdx1* homology models using the interaction generation protocol in DS. Ideally a pharmacophore should not consist of more than 7 features to make computational library searching practical and less resource intensive. The chemical features identified in the *PfPdx1* active site were hierarchically clustered in order to generate a representative pharmacophore hypothesis. Of the chemical features identified, those in close proximity to the ligand (8 Å) were selectively retained during hierarchical clustering. The procedure identified pharmacophore H-bond donors, H-bond acceptors and hydrophobic features (Table 2.7). The individual pharmacophore features for each model are shown as stereo images (Figure 2.15, as well as Figure 2.A5 – Figure 2.A7 in the Appendix 2). The chemical features from the respective pharmacophores included residues known to have interactions with the Ru5P (in *TmPdx1*) and R5P substrates (in *PbPdx1*). Correlating chemical features were identified from the respective pharmacophore hypotheses, such as the side chain of K83 as H-bond donor in all except the custom assigned pharmacophore. A few pharmacophores contained chemical

features associated with residues which were not present in other pharmacophores, such as E107 in the *PfPdx1*_MPD pharmacophore which was flagged as H-bond donor feature, and is absent from other models. This suggested there were five different and diverse pharmacophore hypotheses generated, each representing a different set of chemical features of the R5P active site in *PfPdx1*.

From the *PfPdx1*_MPD structure (Figure 2.15 A) several H-bond donors were identified, and these include the NZ atom of K83, the main chain N-atoms of both E107 and G216. The side chain carbonyl O-atom of D26 and the main chain carbonyl O-atoms of A154 and A214 were identified as H-bond acceptors (Figure 2.15 A). Additionally, two hydrophobic features were identified in the central part of the cavity, which normally contains the ligand, located near E107. In the *PfPdx1*_Ru5P model the K83 as well as D26 was identified as H-bond donor and acceptor features, respectively, and were shared features amongst the models. Additionally side chain N-atoms from R149 as well as main chain N-atoms from G152 and A154 were recognised as H-bond donors.

Table 2.7: Summary of pharmacophore features identified from *PfPdx1* homology models. Specific residues and atoms (in parentheses) involved in either H-bond acceptor or donor interactions as well as residues located near hydrophobic features are indicated. A custom pharmacophore was also built, in which pharmacophore features were manually assigned and aligned to residues in the *PfPdx1*_Ru5P model. Pharmacophores involving water atoms are labelled “(W)”.

Model	Residues identified as H-bond donors (atoms)	Residues identified as H-bond acceptors (atoms)	Residues flagged near hydrophobic features
<i>PfPdx1</i> _MPD	K83 (NZ) E107 (N) G216 (N)	D26 (OD) A154 (O) A214 (O)	Central cavity region, located near E107.
<i>PfPdx1</i> _Ru5P	K83 (NZ) R149 (NE) G152 (N) A154 (N)	D26 (OD) A214 (O)	Central cavity region. Involving residues V108 and A154.
<i>PfPdx1</i> _MPDw	K83 (NZ) G216 (N) F235 - (W) - G216	D26 (OD) I24 - (W) - D26 V108 - (W) - R149	Central cavity region.
<i>PfPdx1</i> _MPDsco	K83 (NZ) A154 (N) G216 (N)	E107 (OE) F235 (O)	Close to V108, near to G152. Second feature near M45 and G237.
Custom assigned pharmacophore	R149 (NE) G155 (N) G237 (N) S238 (N)	D26 (OD) D104 (OD)	-

Consensus atoms involved in H-bond donor interactions identified from the four models include the side chain NZ-atom of K83, as well as main chain N-atoms from both R149 and G216. Consensus atoms predicted as H-bond acceptors include the side chain carboxyl O-atom of D26, and main chain carbonyl O-atoms of A154 and A214 (Table 2.7). Three water molecules in the *PfPdx1*_MPDw model were predicted with pharmacophore features. A water molecule predicted to

H-bond with both F235 and G216 was flagged as an H-bond donor, whereas water molecules located between I24 and D26, as well as between V108 and R149 were flagged with H-bond acceptor attributes (Figure 2.A5). In addition to pharmacophore features identified exclusion constraints were added to restrict mapping of ligands into areas containing amino acids which were not part of the pharmacophore. The exclusion constraints were assigned to all amino acids outside a 5 Å radius of the identified pharmacophore. These exclusion constraints impose geographical restraints which restrict ligands from binding outside the R5P-active site region.

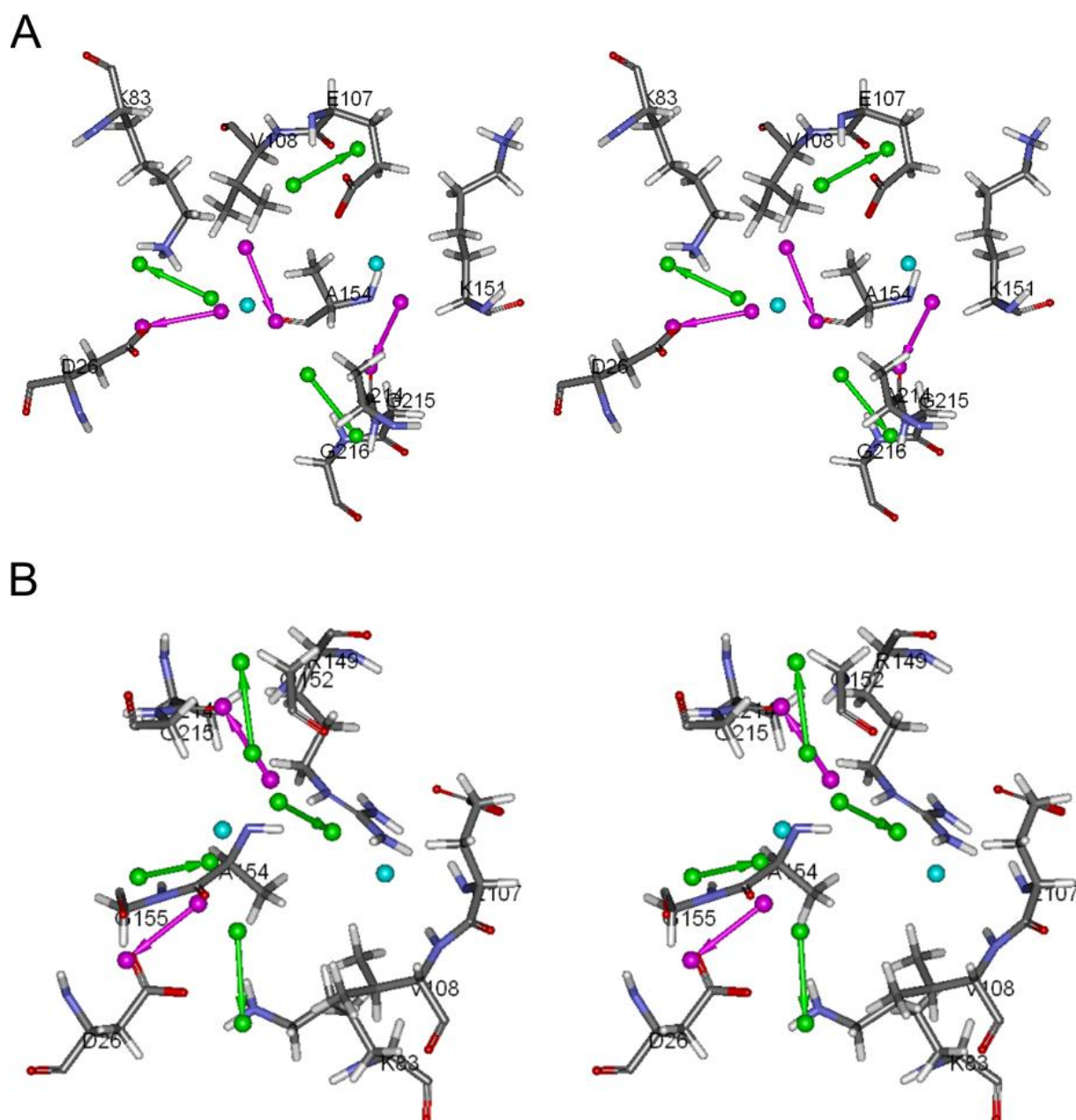


Figure 2.15: Stereo view of the pharmacophores extracted from the *PfPdx1* homology models. Pharmacophores from the **A)** *PfPdx1_MPD* model and the **B)** *PfPdx1_Ru5P* model are shown with chemical features such as H-bond donors (green), H-bond acceptors (purple) and hydrophobic features (cyan) indicated.

2.3.5 Database searching and ligand pharmacophore mapping (LPM)

A database consisting of around 2.1 million commercially available compounds from the Zinc database drug-like subset [174] (zinc.docking.org) were searched using the individual pharmacophores generated from *PfPdx1* models. The drug-like database consisted of ligands (molecules) pre-filtered using Lipinski's rule of five [184]. These rules or physiochemical properties such as; molecular weight ≤ 500 , $\log P \leq 5$, H-bond donors ≤ 5 and H-bond acceptors ≤ 10 , are associated with compounds which have good oral bioavailability or are "drug-like" [185]. The drug-like subset database was previously generated as a multi-conformer composite database using catDB with a maximum of 250 representative conformations (Accelrys Software Inc.) [186]. The drug-like subset database was screened against 5 pharmacophore models and 300 best-hit compounds were identified based on their pharmacophore fit-values. The database searching approach was employed as pre-selection method for compounds that match the pharmacophores. The 7028 ligands identified from the searching procedure (results not shown) were further refined by ranking the pharmacophore fit-values using ligand pharmacophore mapping (LPM). In this procedure, the ligands are strictly aligned to the respective pharmacophore, and a score is assigned for each ligand based upon corresponding distances from pharmacophore features. LPM identified a total of 3063 ligands which satisfied the pharmacophore match constraints from the five pharmacophores (Figure 2.16). The *PfPdx1*_MPD_{sco} model pharmacophore had the most hit ligands, whereas the *PfPdx1*_MPD model pharmacophore had the least (Figure 2.16). A comparison of ligands identified from the models showed that the *PfPdx1*_MPD_w and *PfPdx1*_MPD_{sco} had around 217 shared hits (Figure 2.16). Only one molecule was shared in all of the pharmacophores (Figure 2.16).

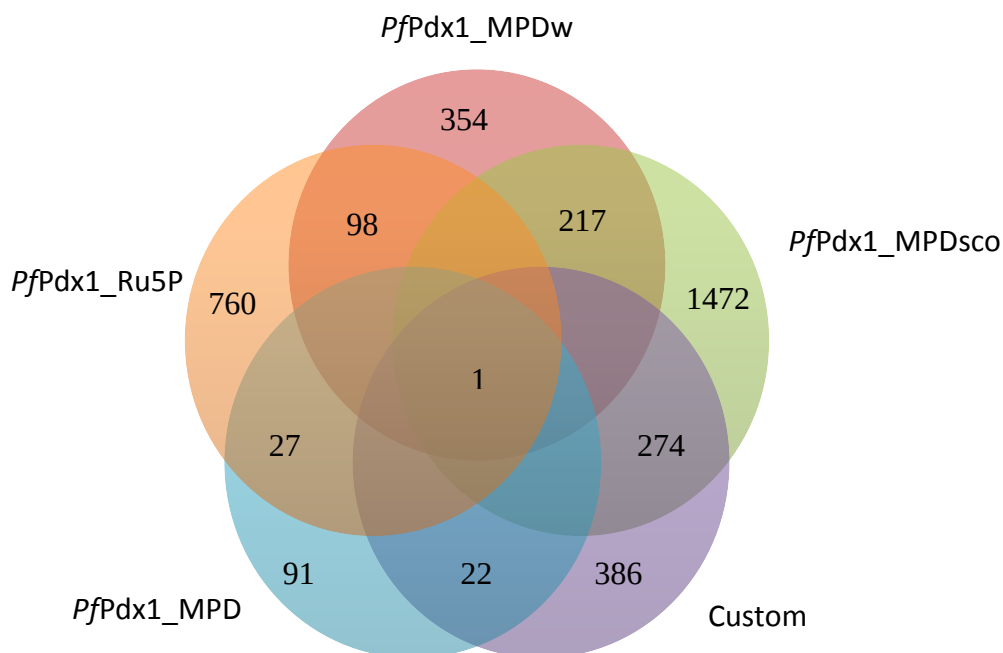


Figure 2.16: Venn diagram representing the number of ligands identified from the respective *PfPdx1* pharmacophores. Common or shared hits are shown in overlapping regions and for simplicity, comparisons were made linearly with adjacent pharmacophores. The ZINC03957744 compound was identified as a hit in all five pharmacophore models.

Compounds were rigorously filtered based on their interaction with the pharmacophore in which at least $n - 1$ corresponding chemical groups had to align to a pharmacophore of n total features. This weighting during pharmacophore mapping ensured that chemical features of compounds that matched pharmacophore features were scored in the same manner. The sulphonamide ligand ZINC03560463 was found to satisfy corresponding pharmacophore features in the *PfPdx1_Ru5P* pharmacophore (Figure 2.17). The sulphonamide ligand ZINC03560463 was predicted to fit in four of the five pharmacophore models, and had the best fit-score in the *PfPdx1_Ru5P* model (Table 2.8). The top 10 best-scoring ligands identified during LPM based on their combined or summed fit values from the individual pharmacophores are shown in Table 2.8. The fit values of the top 100 ligands ranged from 8.45 to 3.66 (complete list not shown).

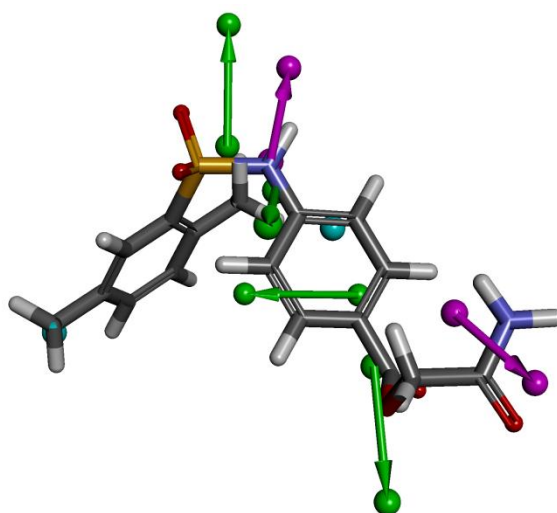
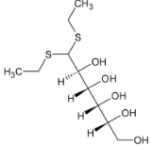
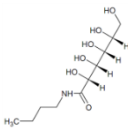
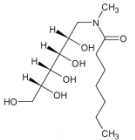
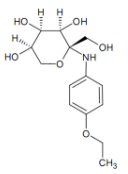
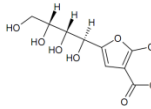
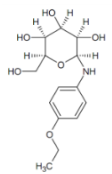
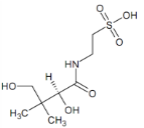
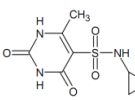
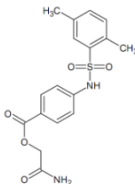
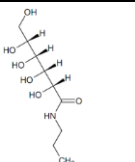


Figure 2.17: Complementary chemical features of ZINC03560463 that matched the *Pf*Pdx1_Ru5P pharmacophore. Only complementary features to the pharmacophore were matched, for example the amine substituent on the right side of the molecule, which is a H-bond donor, matched the H-bond acceptor feature (purple) of the pharmacophore. Similarly, the bottom H-bond donor feature (green) was matched by alignment of the O-atom of the ester group in the molecule.

Chemically diverse ligands were identified, however several of the best scoring ligands shared polyhydroxyl substituent characteristics or carbohydrate sugar backbones (Table 2.8). Compounds that did not match particular pharmacophore features in the respective models had a zero fit score. The representative summed fit-value accounted for mismatches in the pharmacophore fits, and enabled an average representation of ligand matching in the different *Pf*Pdx1 models. Other ligands with comparable fit-values had aromatic ring substituents, with sugar-like moieties (Table 2.8). Of the 3063 ligands ZINC01586791 had the best sum fit-value from all the pharmacophores.

Table 2.8: Best 10 ligands identified using LPM of PfPdx1 pharmacophores. The scores of the ligands from each pharmacophore were combined as a sum fit-value.

Name	Structure	Fit value in pharmacophore					Sum fit-value
		PfPdx1_MPDw	PfPdx1_Ru5P	PfPdx1_MPD	PfPdx1_MPDsco	Custom	
ZINC01586791		3.07	1.70	0.00	0.84	2.84	8.45
ZINC04533078		3.82	0.25	0.00	0.00	3.36	7.43
ZINC04546449		2.42	1.28	0.00	3.02	0.68	7.39
ZINC04773149		2.57	0.00	0.00	1.87	2.07	6.51
ZINC04206966		1.51	0.00	0.00	3.03	1.49	6.04
ZINC04353125		1.33	2.06	0.00	1.65	0.91	5.95
ZINC01666581		2.43	0.00	0.00	0.00	3.17	5.60
ZINC00093092		2.85	0.00	0.00	0.00	2.72	5.57
ZINC03560463		0.55	2.98	0.00	2.00	0.01	5.54
ZINC04533049		3.78	0.92	0.00	0.00	0.77	5.46

ZINC01586791 had the best combined fit score of all the pharmacophores (Figure 2.18 A). The hydroxyl substituents in the ligand were predicted to satisfy corresponding H-bond donor- and acceptor-interactions in the PfPdx1_MPDw (Figure 2.18 A). Similarly, hydroxyl substituents of the ZINC04206966 ligand mapped in the PfPdx1_MPDsco pharmacophore were identified based on both H-bond acceptor and donor properties (Figure 2.18 B). The molecule ranked fifth-best and had a combined fit score of 6.04 of all the pharmacophore models (Table 2.8).

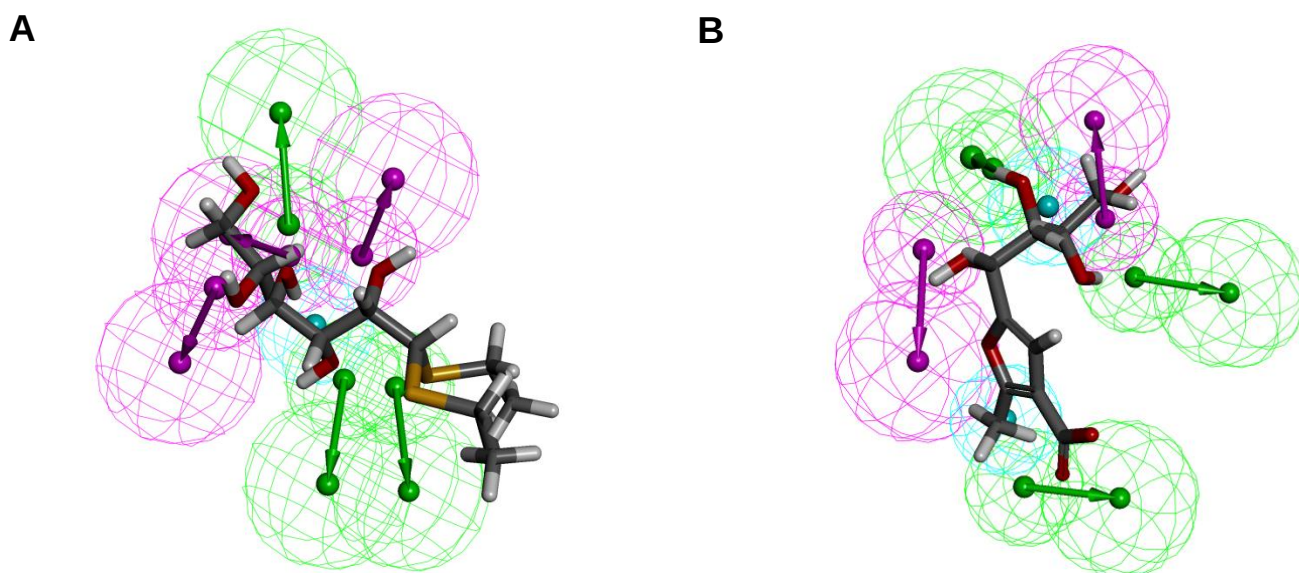


Figure 2.18: Ligands mapped into the *PfPdx1* pharmacophore models using LPM. **A)** ZINC01586791 mapped in the *PfPdx1_MPDw* pharmacophore. The molecule was mapped into the pharmacophore in which 6 corresponding chemical features could be matched out of the 7 pharmacophore features present. A hydrophobic region matching the hydrophobic feature (cyan) was not satisfied. Hydroxyl substituent groups were found to match both H-bond donor (green) and acceptor (purple) features, in which the O-atom could act as an H-bond acceptor in turn matching the respective H-bond acceptor feature area in the pharmacophore, and vice versa for the H-atoms matching H-bond donor feature areas. **B)** ZINC04206966 mapped in the *PfPdx1_MPDsco* pharmacophore. The molecule satisfied 6 of the 7 pharmacophore features, of which the top hydrophobic feature (cyan) was not matched by a corresponding chemical moiety.

Chemical diversity of the 50 best-fitting compounds was assessed by pairwise similarity clustering using the online Chemmine tool. Clustering revealed the best 50 ligands from pharmacophore mapping could be grouped into distinctive groups based on their overall structure (Figure 2.19). Blocked regions of intense red represented molecules with highly similar chemical arrangements, whereas compounds with lesser degrees of similarity are indicated by shades of yellow to white for the least similar compounds (Figure 2.19). The overall distributions of chemical similarity suggested clusters of related compounds were identified from the *PfPdx1* pharmacophores. Numerical values next to the ligand names in Figure 2.19 represent the best-scoring compounds as numerically listed in Table 2.8. The LPM was therefore not randomized and suggested systematic selection based on

the pharmacophore feature matching. A chemical cluster was found to contain compounds with polyhydroxyl characteristics, such as the best scoring compound ZINC01566791 (Table 2.8), and ZINC04533049 (Figure 2.19). This cluster also contained most of the best-scoring ligands from Table 2.8 (Figure 2.19).

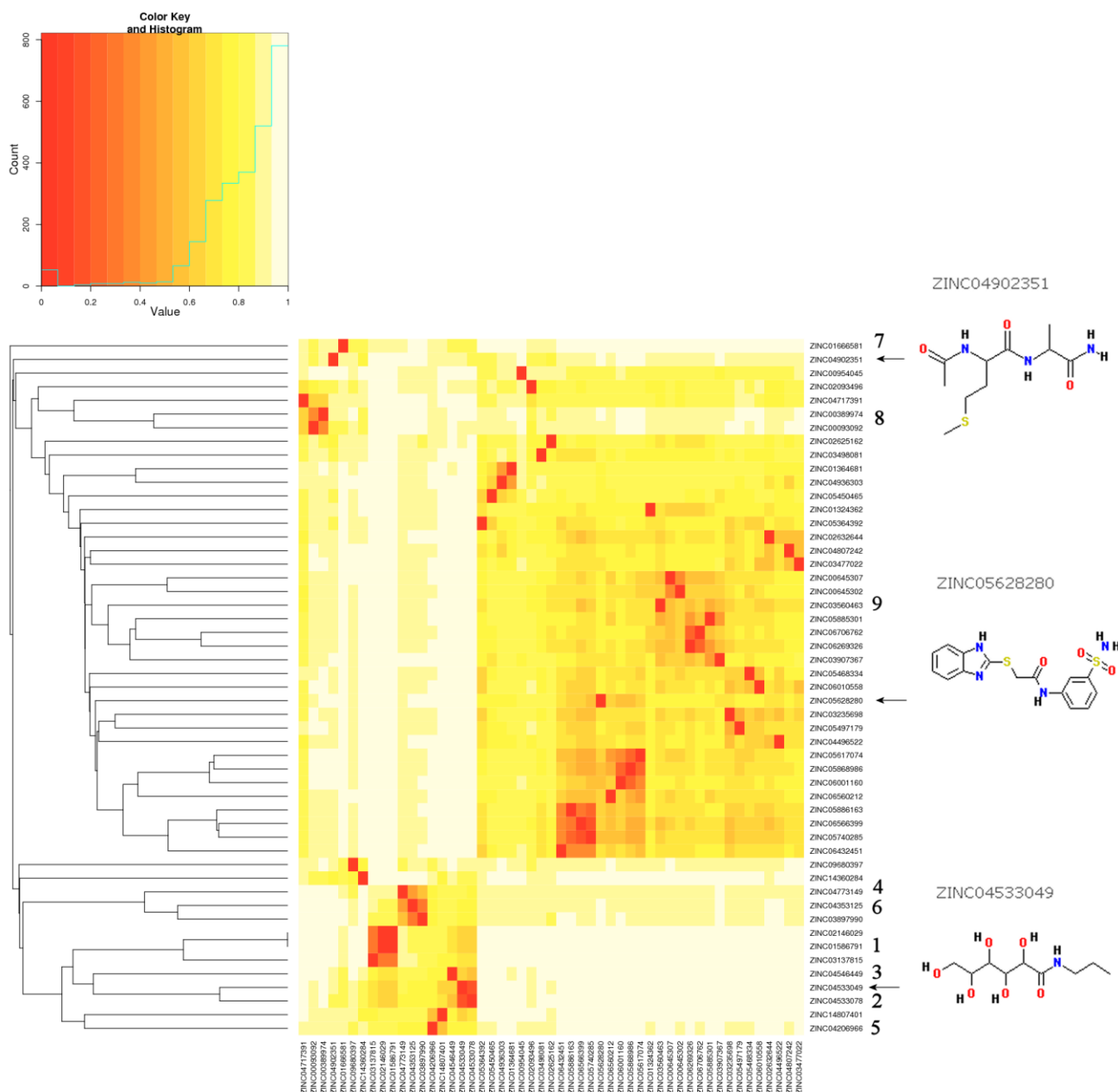


Figure 2.19: Structural pairwise similarity clustering of the best 50 pharmacophore-identified ligands by chemical similarity. Similar ligands are indicated in shades of red, in which the most intense red represents a high degree of chemical similarity or compound identity. Numerical values next to the ligand names, on the right side of the figure, represent the best-scoring in which 1 represent the best scoring ZINC01566791 ligand. The heat map was generated using Chemmine (<http://chemmine.ucr.edu>).

Distinctive scaffolds in the cluster include ZINC04773149 and ZINC04353125 (No. 4 and 6 respectively in Figure 2.19) both of which share aromatic rings connected to a hexose sugar moiety. Molecules such as ZINC01666581 and ZINC00093092 (No.7 and 8 in Figure 2.19, respectively) were clustered together due to similar number of sulphur atoms although still chemically diverse. ZINC00093092 is aromatic compared to the aliphatic peptide-like arrangement in ZINC01666581. Within the 50-best ligand cluster the molecule ZINC04902351 (indicated in Figure 2.19), which had a peptide backbone characteristic, was similar to ZINC01666581 (No. 7 in Figure 2.19), however, as indicated by the heat map, did not share a high degree of chemical similarity to other ligands. Another large cluster of ligands sharing similar aromatic arrangements included the molecule ZINC03560463, with a sulphonamide moiety (No. 9 in Figure 2.19, also previously shown in Figure 2.17), and ZINC05628280, a phenylated-benzoimidazol (Figure 2.18). These ligands had greater aromatic character and larger molecular weights compared to ligands from other clusters. There was also a greater distribution of similarity shared amongst ligands in this cluster as indicated by shades of red in the blocked regions (Figure 2.19).

The average log *P* values (partition coefficient) and molecular weight distribution compared to the summed LPM fit values of the best 277 ligands are shown in Figure 2.20 A and B. The best 277 ligands had summed fit values ranging from 8.5 – 2.6, and ligands with higher summed fit values tended to have smaller molecular weights (Figure 2.20 A) with more hydrophilic character with log *P* values smaller than -0.5 (Figure 2.20 B). This suggested that the LPM approach mostly ranked smaller molecules with greater hydrophilic character as better ligands for the R5P-active site pharmacophore. This correlated with the features in the respective *PfPdx1* pharmacophores, which tended to contain more H-bond donor and acceptor features compared to the hydrophobic features. The higher ranked ligands had physiochemical properties which correlated with the substrate molecule R5P (228.09 g.mol⁻¹ with a log *P* of -3.06). The distribution of ligand molecular weights from the drug-like subset from the Zinc database (before screening) ranged from 150 to 500 g.mol⁻¹, with the log *P* values ranging between 5 to -7 [174]. Within these ranges, the overall distribution of ligand hits from the *PfPdx1* pharmacophore LPM tended to be smaller on average (357 g.mol⁻¹), with slightly hydrophobic character (Avg. log *P* = 1.97). The LPM approach was therefore not only selective for smaller hydrophilic ligands, but able to identify a diverse range of molecules with a wide spectrum of physiochemical properties.

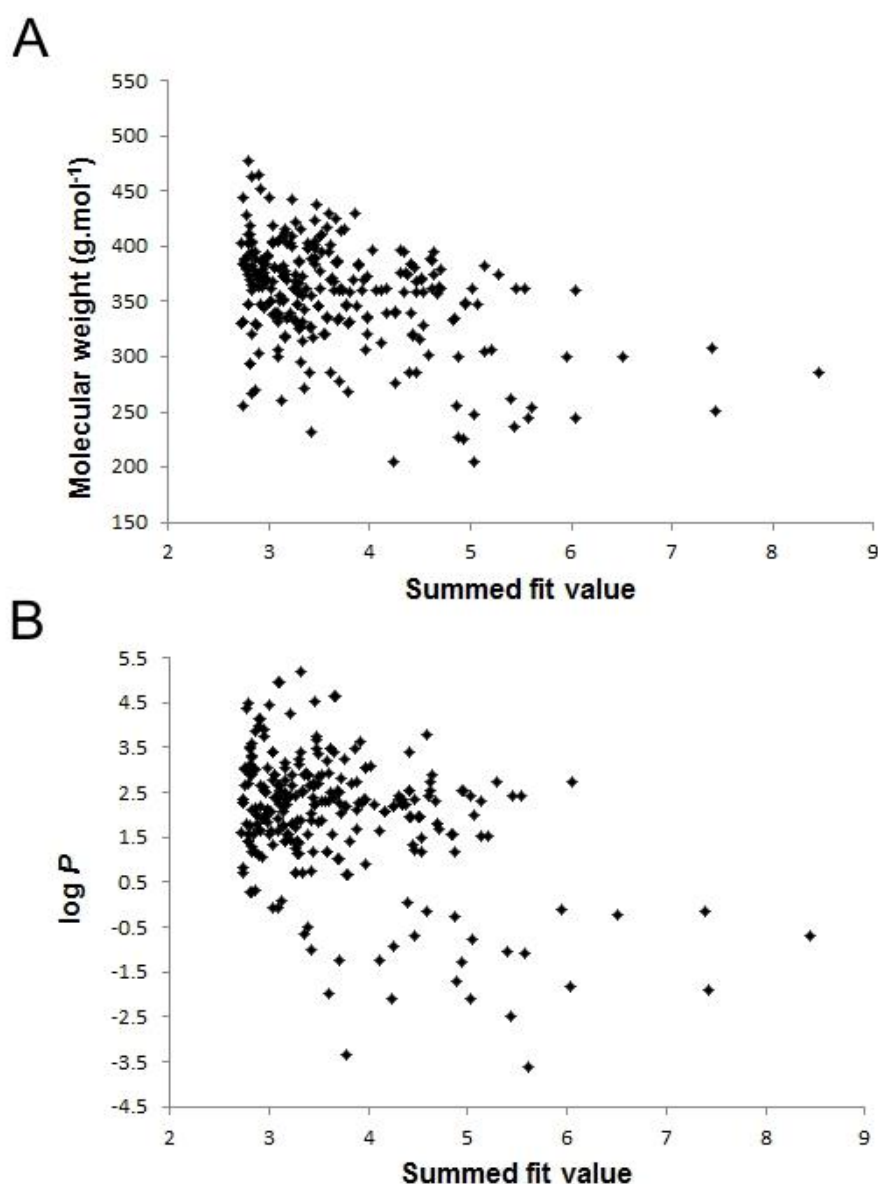


Figure 2.20: The physiochemical properties of the best 277 LPM ligand hits. The $\log P$ and molecular weight distribution of the best 277 LPM hit ligands were plotted against the summed fit values. The average molecular weight was 357 g.mol^{-1} , and the average $\log P$ value was 1.97. Chemical properties were obtained from the ZINC database.

2.3.6 Docking

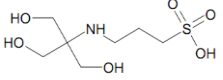
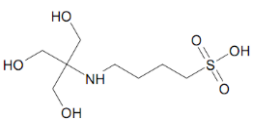
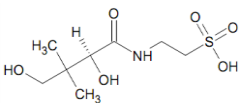
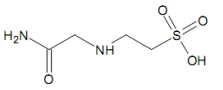
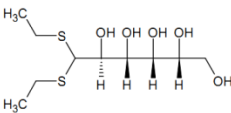
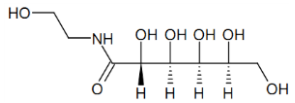
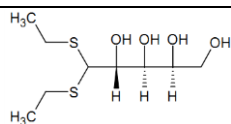
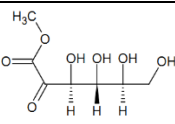
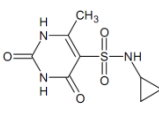
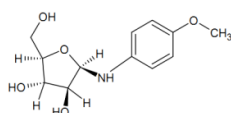
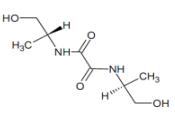
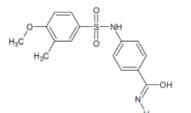
Molecular docking simulations were used to confirm the binding modes of the top 300 ligands, and aided in refining the LPM hits. Ligands were docked into the R5P active site of PfPdx1 homology models, and ligand orientations were scored using Ligscore scoring functions as described in the Methods section. The receptor-ligand binding affinities computed using the LigScore scoring functions have units of pKi ($-\log K_i$) reflecting the predicted association between ligands and receptors.

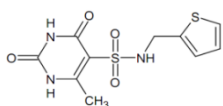
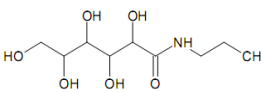
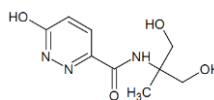
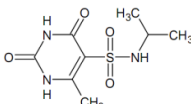
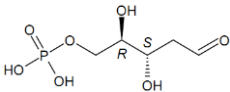
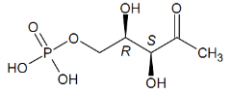
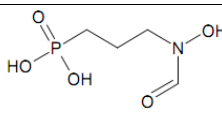
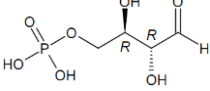
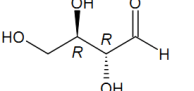
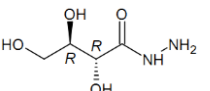
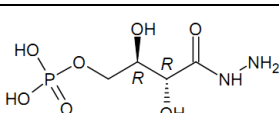
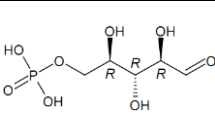
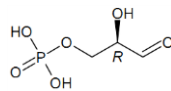
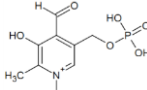
Following the release of the *PbPdx1* structure in 2012 [164], additional docking was performed on the *Pdx1* homologue for selected ligands. Docking into *PfPdx1* supported pharmacophore mapping results, and some of the best scoring compounds from the former approach also had good Ligscores (Table 2.9). Some of the ZINC compounds were not commercially available (or affordable) and necessitated the inclusion of closely related analogues, and their docking values were also calculated in order to support selection for down-stream *in vitro* testing. Analogues of R5P were included and expected to have good active site binding capabilities.

Compounds which docked into the *PfPdx1* R5P active site with large Ligscore values, or binding affinities, comparable to the control R5P ligand included ZINC01666581 (Figure 2.21 A). The sulphonated compound adopted an orientation similar to R5P (results not shown) with the sulphonated group H-bonding to corresponding residues known to stabilise the phosphate group of R5P in *TmPdx1* [157]. This also corresponded with the LPM hit results in which this ligand was ranked 7th best. Several sulphonate analogues of ZINC01666581 were also included during docking to probe the binding effect of various substituted groups containing the sulphonate scaffold. The sulphonate groups of the ZINC05273895, TAPS and ACES were predicted with good ligand binding affinities (Table 2.1), and were orientated in similar positions to phosphate groups of R5P during molecular docking (results not shown). The ZINC05273895 ligand with shorter aliphatic chain had better predicted ligand binding affinities compared to TAPS. The sulphonated analogue ACES had the least substituted chemical groups, however was predicted with the poorest binding affinities compared to the other analogues. The two amino groups on ACES, of which the central secondary substituted group carries a positive charge, resulted in this ligand shifting away from the phosphate binding cavity region to H-bond with D104 and D26 (results not shown).

Other ligands which were predicted with good fit values in the pharmacophores such as ZINC01586791 were shown to have docking scores comparable to R5P (Table 2.9). The ZINC03897990 ligand was predicted with better binding affinities compared to R5P, whereas the LPM combined fit-value was 4.42 – within the top 100 hits (Table 2.9). This emphasised that docking was a useful complementary approach, capable of highlighting certain ligands with intermediate LPM scores. Docking confirmed the feasibility of the LPM approach in being able to identify ligands with good ligands binding affinities. Considerations brought into account during docking include active site shape matching, van der Waals interactions as well as the buried surface area between the ligand and the receptor representing attractive interactions. Some additionally included steric restraints introduced during docking could filter out ligands which are geometrically too large for the R5P active site.

Table 2.9: Ligands from the pharmacophore hits docked into *Pf*Pdx1 models. Rationally selected compounds, which shared similarity to pharmacophore-identified ligands, as well as LPM hits were docked into the *Pf*Pdx1 homology models. R5P was used as reference control during docking. LPM fit values of the ligands are given. LigScores as well as Dock scores for ligands docked into the *Pb*Pdx1 structure are given in parentheses.

Compound name or ZINC ID	LPM fit-value	Structure	LigScore 1 ^a	LigScore2 ^a	Dock score ^b
ZINC05273895	-		6.17	5.6	44.91
TAPS	-		5.75	5.39	61.35
ZINC01666581	5.60		5.97	5.51	54.76
ACES	-		5.22	4.94	82.18
ZINC01586791	8.45		4.76	5.36	24.01
ZINC04683167	-		6.14	5.71	40.09
ZINC03137815	4.88		5.4	6.03	33.31
NSC163924	-		6.21	5.87	45.38
ZINC00093092	5.57		4.55	4.95	44.91
ZINC03897990	4.42		6.87	6.24	49.51
ZINC14360284	5.04		4.97	4.93	32.52
ZINC06269326	4.46		5.6	5.9	34.71

ZINC04717391	4.61		5.27	6.02	42.14
S309737	-		5.65	5.05	41.51
ZINC09680397	4.94		4.85	5.79	42.94
ZINC00389974	5.07		4.5	5.28	41.49
DR5P	-		6.69	6.47	85.64
DXP	-		6.64	6.29	88.92
Fosmidomycin	-		6.27	5.99	91.33
E4P	-		6.40 (3.00)	6.09 (0.41)	85.17 (120.89)
D-erythrose	-		4.74 (3.78)	4.25 (2.15)	42.19 (48.20)
EHZ	-		5.34 (4.07)	5.39 (3.46)	53.48 (46.63)
4PEHz	-		6.55 (3.73)	6.43 (1.46)	98.96 (117.01)
R5P (reference compound)	-		6.55 (4.32)	6.02 (2.63)	90.15 (107.14)
G3P	-		5.89	5.4	61.80
PLP	-		6.51	6.27	87.50

^aLigScore1 and LigScore2, with units of pK_i ($-\log K_i$), refers to predicted receptor-ligand binding affinities.

^bDockscore refers to the unitless rigid body minimisation energy of the final ligand pose calculated during Monte Carlo trials.

Some rationally selected analogues of R5P which were included during docking include DR5P, DXP, E4P and 4PEHz. These ligands had Ligscore binding affinity values similar to R5P (Table 2.9). DR5P had Ligscore values marginally better than R5P, predicting that this molecule had good binding capabilities in the R5P active site. Other R5P analogues that lacked the phosphate group such as D-erythrose and EH₂ had lower score values compared to their phosphorylated counterparts E4P and 4PEHz in *Pf*Pdx1 models, respectively (Table 2.9). Another ligand which had comparable docking scores to R5P in the *Pf*Pdx1 models included 4PEHz (Figure 2.21 B). Molecular docking simulations predicted that the molecule is spatially orientated with phosphate groups in association with G237 and G216. Similar poses were shown for the R5P substrate in the *Tm*Pdx1 crystal structure [157].

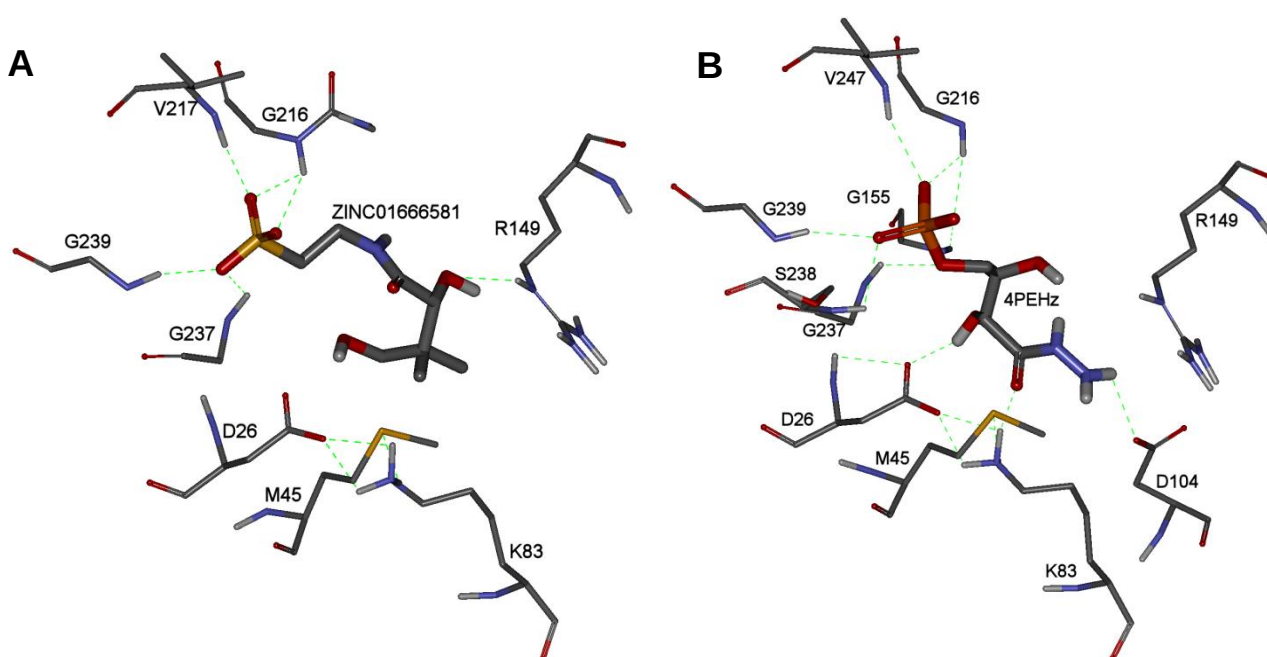


Figure 2.21: Ligands docked in the *Pf*Pdx1_Ru5P homology model. **A)** ZINC01666581 docked in *Pf*Pdx1_Ru5P. The molecule was orientated with sulphate groups H-bonded to G216, G239 and G237, and the C3 hydroxyl substituent was H-bonded to R149. **B)** 4PEHz docked into *Pf*Pdx1_Ru5P predicted that the phosphate group of the molecule to have H-bonds with G155, G216, V247, G239, S238 and G237, whereas the C3 carbonyl O-atom was H-bonded to K83, and the C4 hydroxyl substituent to D26. Atoms C, N, O, P, and H are coloured grey, blue, red, orange and white, respectively.

Additional docking of the ligands in the *Pb*Pdx1 structure was used to confirm predicted ligand binding affinities and orientations as observed in *Pf*Pdx1 models. The R5P-binding site in both *Pf*Pdx1 and *Pb*Pdx1 are highly comparable, and similar binding affinities are expected for ligands in the *Pb*Pdx1 structure. The Ligscore1 receptor-ligand binding affinity of R5P in the *Pb*Pdx1 structure was calculated to be 4.32 (Table 2.9). The reference molecule had similar orientations compared to docking prediction for R5P in the *Pf*Pdx1 models (Figure 2.22 A). R5P analogues, such as E4P and D-erythrose were docked into the R5P-binding site of *Pb*Pdx1 and had marginally

lower Ligscores to compared R5P (Table 2.9). D-erythrose had better predicted ligand binding affinities compared to E4P (Table 2.9). The docked analogue E4P had similar orientations to R5P with the phosphate groups in close association with S239 and G238 (Figure 2.22 B).

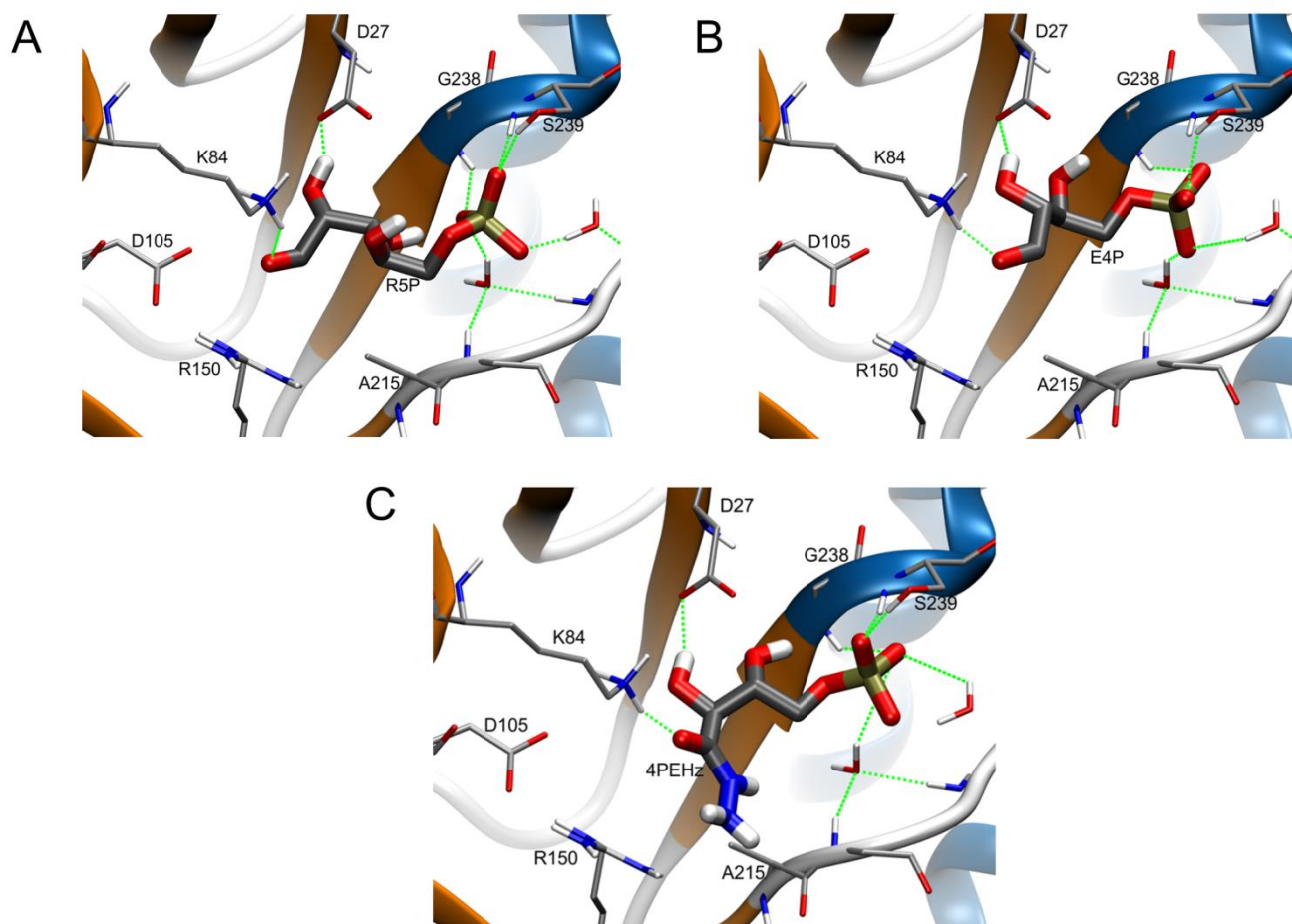


Figure 2.22: Compounds docked into *PbPdx1*. A) R5P docked into *PbPdx1* using LigandFit. The phosphate groups of the molecule H-bonded with G238 and S239, as well as with K84. B) E4P docked into the R5P binding site of *PbPdx1*. The molecule was predicted to H-bond with K84, mainly involving the aldehyde and hydroxyl groups, similarly interacting with residue D27. E4P was calculated to have a LigScore1 of 3.00. C) 4PEHz docked into *PbPdx1*. Similar to R5P, 4PEHz was predicted to adopt an orientation with the phosphate group H-bonded to G238 and S239. The hydrazide group of 4PEHz was orientated approximately parallel to R150 in close proximity to D108 and both ketone and hydroxyl substituents of 4PEHz H-bonded with K84 and D27, respectively.

The 4PEHz ligand had somewhat less favourable predicted interactions in the *PbPdx1* structure, resulting in poorer LigScores, compared to the unphosphorylated EH_z ligand and R5P (Table 2.9). Poses of 4PEHz were scored using LigScore1 and resulted in a value of 3.73, compared to 4.32 for the R5P control. In the *PbPdx1* structure 4PEHz adopted similar positioning compared to R5P, and phosphate groups of 4PEHz were spatially orientated in close proximity to G238 and S239 (Figure 2.22 C). Overall, docking confirmed that compounds from both the LPM and docking approaches

had comparable, if not better, binding affinities to R5P and were suitable candidates for further *in vitro* testing on P/Pdx1.

2.4 Discussion

Homology modelling aims to predict protein conformation based on experimentally determined protein structures and sequence similarities [165]. MSA showed that *Pf*Pdx1 shared around 60% sequence identity with the *Tm*Pdx1 and *Tt*Pdx1 protein sequences. These alignments were used to infer structural information based on sequence conservation to create *Pf*Pdx1 homology models. The level of sequence identity in homology modelling influences the quality of the models. Sequence identities more than 50% are suitable for structure-based drug design and protein-ligand binding predictions [165]. The high degree of sequence identity of *Pf*Pdx1 with *Tm*Pdx1 and *Tt*Pdx1 suggested that accurate models could be generated. The *Tm*Pdx1 structure has a lower resolution (2.90 Å) compared to that of the *Tt*Pdx1 structure (1.65 Å) and at the time this study was initiated *Tm*Pdx1 was the only structure with information regarding the positioning and interactions of the Ru5P substrate. Several *Pf*Pdx1 homology models were built from *Tm*Pdx1 and *Tt*Pdx1 template structures, each incorporating different substrate molecules.

Previous MSA alignments of *Pf*Pdx1 also showed a high degree of identity between plasmodial homologues, all of which contain an additional 20 amino acids on the C-terminal tail end [158]. The Pdx1 homologue from *Toxoplasma gondii* had 58% identity compared to *Pf*Pdx1 [187]. The Pdx1 family of proteins share sequence conservation compared to Dna-K related molecular chaperones – one of the most-conserved protein families [188]. This high degree of conservation emphasises the important cellular and physiological role this protein plays.

The qualities of generated *Pf*Pdx1 homology models were evaluated based on energy violations and spatial restraints. The PDF total energy descriptor represents the summed homology-derived pseudo-energy terms and stereochemical pseudo-energy terms. This is a unit-less value which is useful to interpret the energy violations when individual models are compared. The PDF physical energy is the summed stereochemical pseudo-energy terms which comprise of valence bonds and angles, torsion angles, non-bonded potentials, soft-sphere repulsion energies and improper torsion angles. In both cases higher PDF energy terms indicate larger violations in predicted models. The PDF total energy was also useful to identify and resolve any particular residues with higher restraint violations. These values have no meaning by itself, and more useful to compare different conformers of homology models of the same protein. A lower total PDF energy value determined for *Pf*Pdx1_Ru5P suggested smaller restraint violations in the model incorporating the Ru5P ligand and showed less stereochemical pseudo-energy violations compared to the MPD-containing models.

The DOPE score, a statistical atomic distance-dependent potential descriptor, developed by Shen *et al.*, is a measure of protein stability or native state which can be correlated with model error [179].

The DOPE scores were computed for *PfPdx1* homology models, and were an indication of the conformational energy as well as relative stability of the generated conformation. Considering that the DOPE score does not normalise for the size of the protein, the absolute score is not useful and instead the score is more useful in reporting the relative energies of multiple protein conformations generated during the homology modelling. Homology models with the lowest DOPE scores were selected for further down-stream predictions. The *PfPdx1*_MPD and *PfPdx1*_Ru5P models had similar DOPE scores and therefore comparable conformational energies.

Minimisation geometrically optimizes protein structures and removes any steric overlaps to relax the conformational state of the protein. Ramachadran plots revealed that minimised *PfPdx1* models had less residues in most favoured regions and a transition into additionally allowed regions, as well as disallowed regions. RMSD comparisons average the deviation of atom positioning in compared macromolecules, and give an indication of structural convergence of model-to-template superimpositions. RMSD comparisons of minimised *PfPdx1* homology models suggested a good correlation, or small deviations, to the template Pdx1 structures. The minimised *PfPdx1* homology models also correlated to the *PbPdx1* structure [164]. RMSD calculations of the C α and main chain atoms in the active site regions of *PfPdx1* homology models compared to the *TmPdx1* and *TtPdx1* templates, as well as compared to *PbPdx1*, were close to 1.0 Å. This suggested comparable spatial orientations of R5P-binding site residues and that accurate information could be extracted from the *PfPdx1* R5P-binding site and residues to represent the pharmacophore.

Flicker *et al.* previously modelled the *PfPdx1* using the *BsPdx1* (PDB ID 2NV2) template structure [159]. The model was shown with two additional amino acid residues on the N-terminal. These were proposed to add additional H-bond character on the essential N-terminal component which interacts with the Pdx2 monomers. In this study *PfPdx1* homology models could be created with similar N-terminal α -helical arrangement compared to *TmPdx1* and *PbPdx1*. Residues in the α N region of *PfPdx1*, which shared approximately 70% identity to the *PbPdx1* structure, could be superimposed with the *PbPdx1* structure. Residues in the *PfPdx1* N-terminal region are expected to have similar interactions with *PfPdx2* compared to the interactions of *PbPdx1* with *PfPdx2* [164].

The topology of the final minimised protein homology model was comparable to the *TmPdx1* crystal structure (PDB ID: 2ISS). The C-terminal tail of *PfPdx1*, consisting of 20 amino acids, included in homology models did not adopt any secondary structural conformations and remained unstructured. In the *TmPdx1* structure only a short part (between residues 270 – 280) was shown to become ordered when Pdx1 dimerises with Pdx2, and the furthest terminal end residues remained disordered [157]. The *PbPdx1* structure, in which the last 10 residues were not present, showed that

the last 30 amino acid residues of this C-terminal tail did not adopt any secondary conformation when dimerised with *Pf*Pdx2 (chimera with *P. falciparum* Pdx2 counterpart) [164]. The C-terminal region of *Pf*Pdx1 models was similarly disordered, and could be aligned with *Pb*Pdx1 residues up until residue T281.

It was shown that the *Pf*Pdx1 $\Delta_{270-301}$ and *Pf*Pdx1 $\Delta_{273-301}$ C-terminal deletion mutants were incapable of binding R5P and catalytically inactive, whereas a shorter mutant *Pf*Pdx1 $\Delta_{279-301}$ was capable of binding R5P and forming PLP [182]. These studies suggested that residues preceding G279 were important, and could have communicated long range interactions to the R5P-binding site. C-terminal residues of the *Pf*Pdx1 models had spatially similar arrangement to those found in the *Pb*Pdx1 structure, which suggested accurate representation of residues that could be involved in long range contacts with residue of the R5P-binding site.

Consensus R5P-binding site residues in *Pf*Pdx1 models and templates included the R5P-binding residue K83 as well as D26, which were predicted to form H-bonds with the R5P substrate. Some noticeable differences between the *Tm*Pdx1 and *Pf*Pdx1 active site could include the hydrophobic contribution of the M44 residue. This residue was shown to have long range hydrophobic interactions with the C3 and C2 atoms of Ru5P in *Tm*Pdx1. In the *Pb*Pdx1 structure the equivalent M46 residue was as in close proximity to R5P, which was covalently bonded to the K84 residue [164]. In minimised *Pf*Pdx1 models the equivalent M45 was spatially orientated with sulphur groups closer towards the R5P substrate. The methyl group however occupied a similar positioning compared to other Pdx1 structures. The hydrophobic contributions of this residue are expected to be similar compared to *Tm*Pdx1 and *Pb*Pdx1. This residue was identified to form part of a hydrophobic pharmacophore feature shared with the G237 residue in the *Pf*Pdx1_MPDsco model. Other residues which contribute hydrophobic features in the *Pf*Pdx1 pharmacophore included V108, which is also implicated in *Tm*Pdx1 (V107 in *Tm*Pdx1) [157].

The pharmacophores extracted from the *Pf*Pdx1 homology models were diverse, with different residues in the R5P-binding site contributing to the final pharmacophore model. The R5P-binding site residue K83 plays a fundamental role in initial imine formation with either Ru5P or R5P [157, 164]. This residue was consistently identified as an H-bond donor in all pharmacophores, confirming that representative pharmacophores containing important features could be created. Considering that imine formation entails nucleophilic attack of the amine on the carbonyl group of R5P, K83 can become deprotonated during the course of the reaction coordinate. Nonetheless, in either the protonated or deprotonated states K83 remains a H-bond donor, as predicted in *Pf*Pdx1 pharmacophores. Other residues represented in the *Pf*Pdx1 pharmacophores include D26, of which

the side chain carbonyl group was predicted as a H-bond acceptor feature, capable of sharing the lone pair of electrons on the oxygen atom. This is consistent with the role attributed to this residue in *TmPdx1* as well as *PbPdx1* [157, 164]. Residues A154, G216, G237 and S238 were also identified as predominantly H-bond donor features. These residues are involved in R5P-phosphate group H-bond interactions in *TmPdx1* and similarly in *PfPdx1* can be considered as the phosphate-group pharmacophore region.

Pharmacophore-based screening and confirmatory docking of positive hits was performed using the ZINC database containing 2 million compounds and resulted in the identification of around 3000 ligands. Plotting of the physiochemical properties of the ligands suggested that a chemically diverse set of ligands were identified, but also emphasised that smaller molecules with a hydrophilic character were better matched to the *PfPdx1* pharmacophore. The diversity of the ligands suggested that matching chemical features on the ligands were found, whereas the basic chemical framework these were situated on was different. This becomes useful when active molecules are identified, and the solubility or pharmacokinetic properties need to be optimised. The average molecular weight and log *P* values of the top scoring compounds from the LPM were in close agreement to properties of most approved drug libraries (MW of 375 Da, log *P* of 3.4) [99]. The distribution of physiochemical properties from the initial drug-like library suggested that pharmacophore screening effectively filtered compounds from the database. Clustering of the scaffold composition of the best hits further emphasised the identification of molecules from other diverse chemical scaffolds, and were grouped into distinct classes. The 4 best scoring ligands had similar chemical arrangements, with polyhydroxylated moieties, implying that from the large *in silico* library some selectivity was achieved, and shared scaffolds could reproducibly be identified. The top scoring molecules had modified polyhydroxylated or carbohydrate backbones similar to R5P. pharmacophoreP mapping therefore successfully identified molecules with shared characteristics to known substrates of *PfPdx1*.

In this study docking was used to reinforce pharmacophore hits and introduced additional filtering for selection of ligands for *in vitro* testing. Docking is more computationally intensive, and therefore only the top scoring compounds from LPM were docked. LPM spatially aligns flexible ligands in order to maximise the number of geometrically overlapping chemical features with that of the target or receptor [189]. Docking predicts and generates binding modes of ligands which is comparable to crystallographic ligand binding modes [190]. Docking considers the geometrical fit of ligands in the active site, by calculating the local energy minimum of the complex, whilst incorporating protein flexibility. Additional shape restraints during docking introduced more stringent fitting conditions for further refining of pharmacophore hits. The predicted ligand binding

affinities from docking simulations were used as criteria to evaluate ligand suitability, and helped in assessing ligand binding modes. R5P could be docked back into the *Pf*Pdx1 models and had binding modes comparable to that of Ru5P in *Tm*Pdx1 and R5P in *Pb*Pdx1 [157, 164]. The top scoring molecules from LPM scored were similarly predicted with the high ligand binding affinities during docking.

The LPM fit-values and data from molecular docking were as criteria to select molecules for further *in vitro* testing. Ligands with appreciable LPM fit-values such ZINC04546449 were excluded during docking. Docking in this case allowed the exclusion of ligands which did not match the shape restraints of the R5P active site. The ZINC04546449 ligand was in the top 10 LPM hits, suggesting a good fit in all of the pharmacophores, however contained a C₆ aliphatic portion which was not expected to be accommodated in the active site. The ligand failed to dock in any of the *Pf*Pdx1 models, and confirmed that the ligand should be excluded for further testing on the enzyme. This molecule additionally shared chemical similarity to ZINC04683167, which docked better, and was a better alternative to test.

A few compounds with good LPM fit-values such as ZINC04533078 were not commercially obtainable, and an alternative analogue ZINC04683167 was used. Docking in this case was used to confirm ligand binding affinities, which were within the comparable range of R5P for the selected analogue. Docking was useful to prioritise selection of certain ligands such as ZINC03897990 which had an intermediate LPM fit-score (4.42), ranking 50th on the LPM-hit list, whereas the ligand had predicted binding affinities better than R5P. Docking therefore confirmed the binding capabilities of the ligand. Docking confirmed that ligands with polyhydroxylated moieties such as ZINC01586791 and ZINC03137815 identified during LPM had potential for binding in the R5P active site of the *Pf*Pdx1 models. Analogues of these such as NSC163924 and ZINC04683167 (which replaced ZINC04546449), which were not screened in LPM, were included in the selection for *in vitro* testing to determine the effect of alternative side chain moieties from the polyhydroxylated scaffold.

Molecular docking simulations predicted R5P analogues, such as DR5P, DXP and E4P with better or comparable fit values to R5P in both *Pf*Pdx1 models and the *Pb*Pdx1 structure. The phosphate groups of most of these analogues were predicted to form H-bonds with residues A154, G216, G237 and S238 – the phosphate-binding region in the R5P active site. These corresponding residues in *Tm*Pdx1, which are highly conserved, form H-bonds with the phosphate group of R5P [157]. This emphasized that coordinated anchoring of the phosphate moiety could be an important requirement for binding in the R5P site of *Pf*Pdx1. These analogues were therefore included for

additional testing on *Pf*Pdx1. As linear forms (acyclic) these analogues had either terminal aldehyde or carbonyl groups, a feature shared with the R5P substrate. These groups are considered to be relatively reactive and have the potential to interact with other chemical groups [191]. For this reason these are generally undesirable to include in virtual libraries, and were manually included during docking studies. When considering the substrates of *Pf*Pdx1, G3P and R5P, these molecules bearing the same terminal aldehyde arrangement are also reactive, in particular the reactivity of R5P enables this molecule to bind to K83 in the active site [160]. These shared features on the R5P analogues, such as DR5P and E4P, were therefore appropriate and believed to have potential in affecting R5P binding in *Pf*Pdx1. Additional details regarding the reactivity of R5P, and the binding mechanism in Pdx1 proteins, is discussed in Chapter 3.

Several sulphonated ligands such as ZINC01666581 were identified during LPM, and the binding modes during docking suggested that the sulphonate groups could associate with the components of the phosphate-binding region of the R5P-active site. The LPM and docking therefore suggested that compounds carrying this sulphonate moiety may be accommodated within the R5P-active site, in particular with these groups corresponding to the phosphate group of R5P. It is not uncommon for phosphate groups in ligands to be substituted for sulphonate groups which reduces potential chemical and enzymatic degradation and thus increases stability of compounds [192]. Moreover sulphonate groups carry less negative charge and provide a neutral characteristic to the compound, which is more desirable in order to cross cellular membranes [192]. For this reason additional ligands such as ZINC05273895, TAPS and ACES were included during *in vitro* testing to probe whether sulphonated ligands may be accommodated.

2.5 Conclusions

The current rate at which the limited antimalarial chemotypes are becoming ineffective due to drug-resistance necessitates the discovery of better and more effective chemotherapeutics. Newer therapeutic targets in essential pathways need to be investigated and exploited. PLP biosynthesis in the parasites provides the PLP cofactor to myriad array of different parasite-essential processes, and for this reason has become an attractive pathway for the development of newer chemotherapeutics. Both *PfPdx1* and *PfPdx2* are integral parts of the PLP synthase complex and *PfPdx1* forms the catalytic centre in which all the substrates condense to form PLP.

Homology models created of the *PfPdx1* were comparable to a recently solved *PbPdx1* structure, as well as *TmPdx1*. The high degree of sequence identity and structural homology shared amongst Pdx1 proteins is indicative of structural conservation owed to its specialised biochemical function. When this study was initiated the only definitive substrate binding site in the Pdx1 protein was the R5P-active site. Residues in the R5P-binding site of *PfPdx1* were conserved and had similar spatially orientations compared to *TmPdx1* and *PbPdx1*. Residues previously implicated in substrate molecular interactions were highlighted during extraction of the *PfPdx1* pharmacophores, and were used to create a spatially and chemically representative site of the R5P cavity in *PfPdx1*. Ligands that matched the pharmacophore features were identified, and molecular docking confirmed favourable shape complementary and the potential to form H-bonds with R5P site residues. The *in silico* derived ligands had diverse chemical scaffolds, which suggested that the computational screening effectively searched the initial compound library, and spatially arranged complementary chemical features on the ligands could be found on more than one type of chemical framework. This suggests that the expansion into different lead molecules or scaffold hopping from actives could be possible, an approach often used to improve pharmacokinetic properties and solubility of chemical leads [193].

Docking reinforced LPM and showed that top scoring ligands from pharmacophore screening had favourable shape complementarity and the potential to form H-bonds in the R5P active site. This ultimately aided in selection of ligands for *in vitro* testing. A series of sulphonated analogues were selected for further testing on the protein based on predictions that the sulphonate groups interact with residues in the phosphate-binding region in the R5P-cavity. These may therefore interfere with R5P-binding in the protein, and could be potential inhibitors. Additionally, analogues of R5P such as DR5P and E4P were predicted with favourable binding affinities, and had binding modes comparable to R5P. Several polyhydroxylated sugar-like ligands were also highlighted during docking and were selected for additional testing.