University of Venda

The Search for Novel Compounds Targeting *Pf*CDPK4 for Therapeutic Treatment of Malaria

By

Thomas Makungo

Submitted in fulfilment of the requirements for the Degree of

Master of Science (Chemistry)

in the Department of Chemistry

School of Mathematical and Natural Sciences

at the University of Venda

Supervisor

Prof. T van Ree (Univen)

Co-Supervisors

Dr Tsepo Tsekoa (CSIR) and

Dr Dalu Mancama (CSIR)

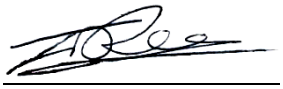
Date submitted: May 2015

Declaration

I, Thomas Makungo (Student No: 11511833), hereby declare that the dissertation for the Master of Science (Chemistry) degree at the University of Venda, hereby submitted by me, has not been submitted previously for a degree at this or any other university, that it is my own, work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signature:  Date: 07/05/2015

Thomas Makungo

Supervisor:  Date: 07/05/2015

Prof. T van Ree

Co-Supervisors:  Date: 07/05/2015

Dr Tsepo Tsekoa

 Date: 07/05/2015

Dr Dalu Mancama

Plagiarism declaration

I, Thomas Makungo (Student No: 11511833), hereby declare the following:

- I am aware and acknowledge that plagiarism is wrong. Plagiarism is the use of another person's ideas or published work and to pretend that it is one's own and it is a serious academic offence for which the university may impose severe penalties.
- I declare that this thesis is my own work and that I have correctly acknowledged the work of others. The work has not been previously submitted to the University of Venda or elsewhere for this degree.
- I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work.

Signature:  Date: 07/05/2015

Abstract

Due to the increasing incidence of *Plasmodium* strains that are resistant to current frontline anti-malarial drugs, malaria remains a global public health challenge. In recent years, the emergence of resistance to frontline antimalarial drugs including the more recently discovered artemisinin class drugs has become one of the greatest challenges of controlling malaria incidence and mortality. There is, therefore, an urgent need to develop novel targets and antimalarial drugs that are effective against drug-resistant malarial parasites. Recent studies have demonstrated that calcium dependent protein kinases (CDPKs) regulate a variety of biological processes in the malaria parasite *Plasmodium falciparum* and that CDPK4 is important for parasite development. The gene disruption of CDPK4 in *Plasmodium berghei*, which results in major defects in sexual differentiation of the parasite has highlighted the importance of CDPK4 in *Plasmodium* biology and suggests that it may be used as a target for therapeutic drugs. *Pf*CDPK4 is expressed in the gamete/gametocyte stage, and this could make *Pf*CDPK4 an essential target for malaria drug discovery. The structure of *Pf*CDPK4 was used as a template in the discovery of malaria drug leads and in designing chemical compounds or inhibitors that will show anti-parasitic activity against the target molecule. The model structure of *Pf*CDPK4 was generated through homology modelling, and model structure validation confirmed that the model structure of *Pf*CDPK4 is of stereochemical quality. The molecular modelling approach of *in silico* screening was utilized in this research, wherein a large library of chemical compounds, some natural chemical compounds, and clinically approved kinase inhibitors were screened against the target molecule *Pf*CDPK4. *In silico* screening of the Bio-Focus library against *Pf*CDPK4 resulted in twenty-six compounds being identified; *in vitro* single screening at a concentration of 5 μM confirmed that three compounds exhibit moderate antimalarial activity against the NF54 strain of *Plasmodium falciparum*, with the percentage inhibition ranging between 42% and 47%.

Keywords: Malaria, antimalarial, molecular modelling, docking, protein kinase, Plasmodium.

Acknowledgements

- I would like to thank Prof T. van Ree for his patient and wonderful supervision. I always enjoy working under his supervision.
- Dr T. Tsekoa for introducing me to the research topic, for his encouragement when I lost my footing and for his wise guidance throughout the studies.
- Dr D. Mancama for his generosity in allowing me to use the Bio-Focus library for my studies and for his co-workers who were responsible for *in vitro* screening for prioritized compounds.
- The National Research Foundation (NRF) for much needed financial support that helped my studies to be successful.
- The Council for Scientific and Industrial Research (CSIR-Biosciences) for letting me use their resources for my studies.
- Centre for High Performance Computing (CHPC) for letting me access the BIOVIA (former Accelrys) Discovery Studio licence via their server.
- The University of Venda for giving financial assistance and the opportunity to acquire the MSc degree under its auspices.

Dedication

I would like to dedicate this dissertation to my family for their love, support and understanding when I am not there for them because of my studies. A special feeling of gratitude to my lovely mother, Tshinakaho Makungo, for her encouragement and believing that my studies will be successful.

List of Abbreviations

ACT	-	Artemisinin-based combination therapy
ATP	-	Adenosine triphosphate
BKI	-	Bumped kinase inhibitor
BLAST	-	Basic Local Alignment Search Tool
CDK	-	Cyclin-dependent kinase
CDPK	-	Calcium-dependent protein kinase
CHARMM	-	Chemistry at HARvard Molecular Mechanics
CQ	-	Chloroquine
ePK	-	Eukaryotic Protein Kinase
FDA	-	Food and Drug Administration
FIKK	-	Phenylalanine-isoleucine-lysine-lysine
GOLD	-	Genetic Optimization for Ligand Docking
HPLC	-	High-performance liquid chromatography
HTS	-	High-Throughput Screening
IR	-	Infra-red
MAPK	-	Mitogen-activated protein kinase
NMR	-	Nuclear magnetic resonance
NSC	-	Natural and Synthetic Compounds
PDB	-	Protein Data Bank
PSI-BLAST	-	Position specific iterated BLAST
SBDD	-	Structure-based drug design
SP	-	Sulphadoxine-pyrimethamine
WHO	-	World Health Organization

Table of Contents

List of Figures in Text

List of Tables in Text

CHAPTER 1

INTRODUCTION

Due to the increasing incidence of *Plasmodium* strains that are resistant to current frontline antimalarial drugs, malaria remains a global public health challenge. There is therefore an urgent need to develop novel targets and antimalarial drugs that are effective against drug-resistant malarial parasites.

Malaria is caused by protozoan Apicomplexan parasites of the genus *Plasmodium*. This disease is transmitted by infected female *Anopheles* mosquitoes [3]. Five malaria parasites infect and cause disease in humans. These are *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*; among them *Plasmodium falciparum* is the most deadly [4-7]. The World Malaria Report published by the World Health Organisation reported that there were about 207 million cases of malaria in 2012 around the world. This resulted in the death of about 627 000 people around the world, with the highest mortality observed on the African continent [4]. However, according to an independent systematic analysis done from 1980 to 2010, mortality due to malaria is estimated to be twice as high as formerly reported, when cases that go undiagnosed and untreated are included. Most deaths occur in western, eastern, and central Africa [8]. It was also estimated that 3.3 million people were at risk of malaria infection in 2011; most of them were again, in Sub-Saharan Africa, with children under five years and pregnant women being most severely affected [9].

1.1 Motivation

1.1.1 Plasmodium life cycle and disease aetiology

The *Plasmodium falciparum* life cycle is well studied and has been comprehensively described by various sources [1,3,5,10,11]. Humans and female *Anopheles* mosquitoes are the two Most important host organisms of the malaria parasite [12]. In humans, infection begins with the bite of an infected female *Anopheles* mosquito, which injects sporozoites into the bloodstream [13,14]. The sporozoites enter the liver within thirty minutes and infect the hepatocytes where they undergo an intense mitotic replication generating several thousand merozoites [15] (see Figure 1). Drugs that target the liver stage are very important to prevent disease development

and to provide a radical cure for *P. vivax* and *P. ovale*, which can hide in the liver as hypnozoites and cause relapsing malaria after some months or even years after the initial infection [1,15,16]. The merozoites invade erythrocytes, where they also undergo schizogony (development of schizont), the process that is responsible for malaria pathogenesis [15]. Pathogenesis only occurs during the erythrocytic cycle [17]. When the liver stage is complete, each liver schizont releases thousands of merozoites into the bloodstream infecting the red blood cells. In the red blood cells the merozoites differentiate into male and female gametes (asexual replication), developing through ring stages consisting of the trophozoite and schizont stages [10]. During the asexual stages, if another female mosquito bites an infected human, the malaria gametocyte is transmitted back to the mosquito host. In the mosquito gut, the male and female gametocytes fuse into a zygote. The resulting zygote forms an ookinete, which penetrates across the mosquito gut to form oocysts. Oocysts rupture and release the sporozoites that travel to the mosquito salivary glands to complete a life cycle. The sporozoites can be injected into humans by the mosquito's next blood meal [1,5]. Health complications that are frequently seen because of *P. falciparum* infection include severe anaemia, cerebral malaria, renal failure and pulmonary infection, leading to death. There are two main features that separate *P. falciparum* malaria from the other human malarias: (1) the ability to invade the erythrocytes, causing a high increase in malaria parasite populations, and (2) its ability to adhere to the venular endothelium of erythrocytes infected with maturing parasites [18].

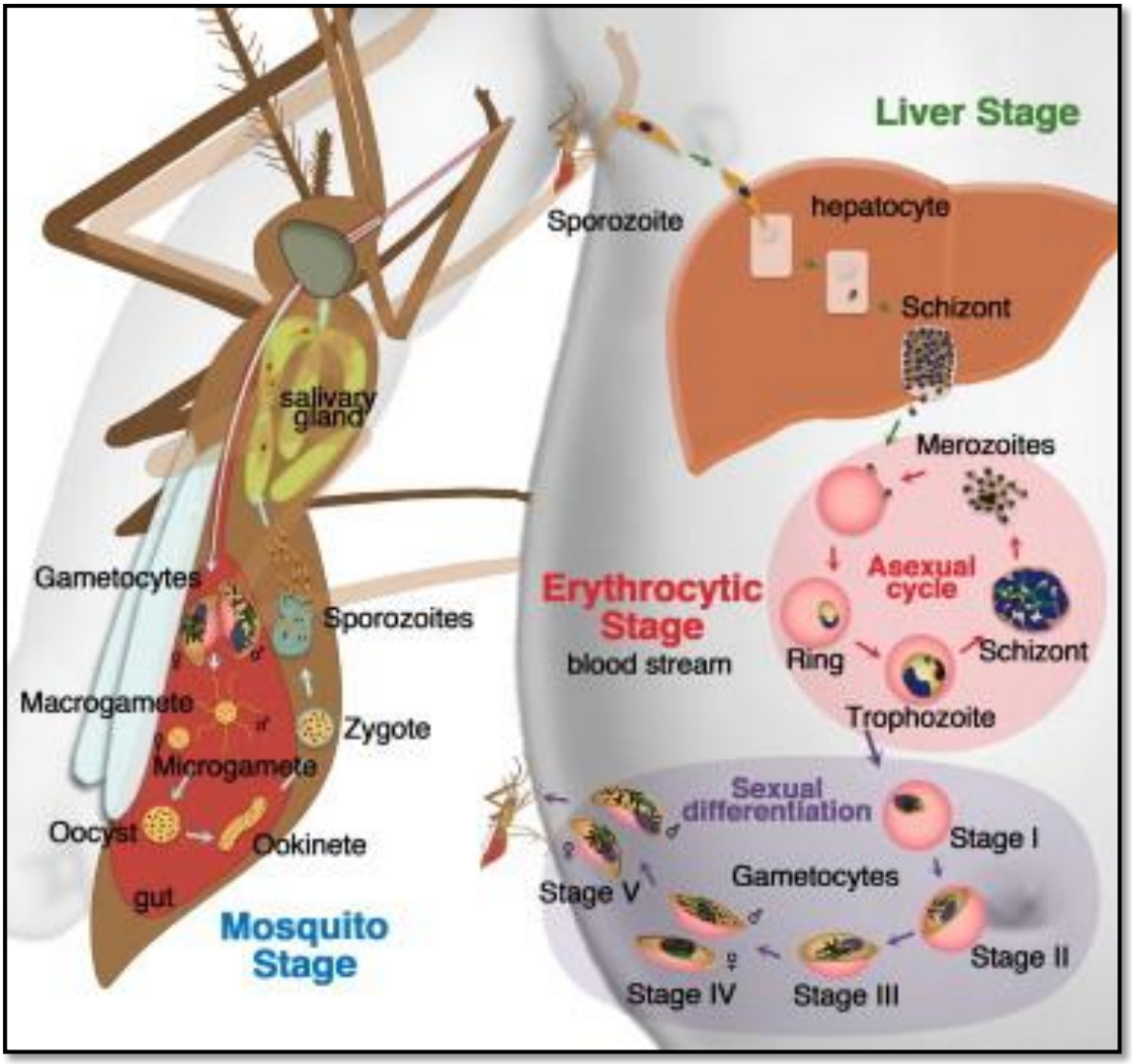


Figure 1: *Plasmodium falciparum* life cycle, taken from [1].

1.1.2 Malaria treatment and resistance

In recent years, the emergence of resistance to frontline antimalarial drugs has become one of the greatest challenges of controlling malaria incidence and mortality due to the disease. Commonly used antimalarial drugs include chloroquine (CQ), sulphadoxine-pyrimethamine (SP), quinine, mefloquine and artemisinin [7]. CQ, SP and artemisinin have been widely used as frontline drugs because they are highly efficacious against *P. falciparum* infected erythrocytes and all patients can use them [6]. However, *P. falciparum* resistance to CQ, mefloquine and SP first emerged in the Cambodia-Thailand border region and resistance then spread from the Greater Mekong sub-region to Africa. Resistance to artemisinin was then identified and confirmed in Cambodia and Thailand in studies conducted from 2001 to 2009

[19], leading to agreement on a global plan for artemisinin resistance containment (GPARC) [20]. Some of the known antimalarial drug structures are shown in Figure 2.

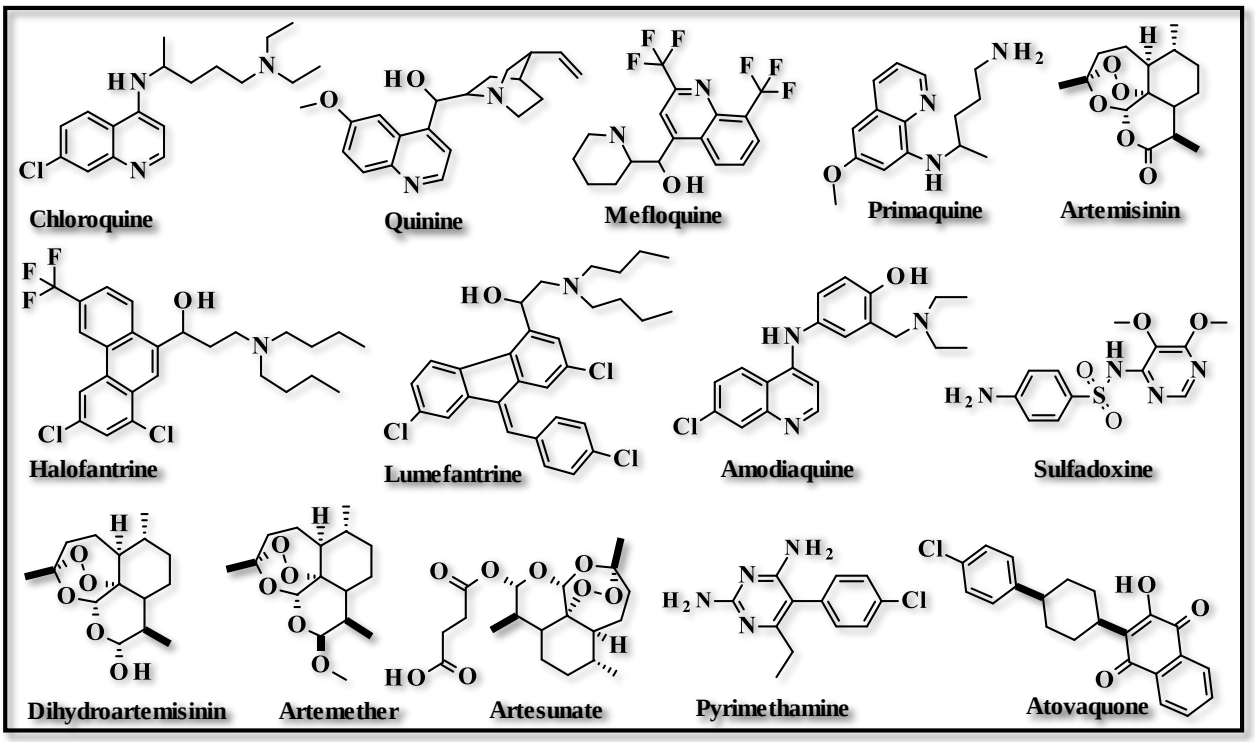


Figure 2: Some known antimalarial drugs [21].

1.1.3 Protein kinases

Protein kinases are enzymes that play the important role of transferring phosphate groups from ATP to specific amino acid residues on protein targets or substrates. This affects the activity, stability and interaction of the target protein with its ligands [22]. The central role played by protein kinases in cellular signalling and the presence of ATP binding pockets that can be accessed and inhibited by chemical compounds has placed protein kinases among the most important and promising classes of therapeutic targets [23].

Recent studies have demonstrated that calcium dependent protein kinases (CDPKs) are directly regulated by calcium and regulate a variety of biological processes in the malaria parasite. Calcium dependent protein kinases (CDPKs) are present in the malaria parasite, *P. falciparum*, and have been found only in plants and some protists [24]. These enzymes and their close homologues are completely absent in humans. There are five members of the CDPK group in

Plasmodium that are present as multi-genes, and it has been proposed that different CDPKs function at different stages of the malaria parasite cycle [25].

Genome analysis identified 86 to 99 genes encoding putative protein kinases in *P. falciparum*, comprising 65 that are related to the family of eukaryotic protein kinase (ePK) as well as the novel FIKK (phenylalanine-isoleucine-lysine-lysine) family with 21 members. Most of the 65 ePKs of *P. falciparum* can be phylogenetically assigned to major clusters or groups that include the AGC (calcium-phospholipid-dependent kinases), CMGC, CK1 (casein kinase 1), CaMK (calcium/ calmodulin-dependent kinases) and tyrosine kinase-like (TKL) groups observed in the human kinome [26]. The CMGC group is one of the major ePK groups, comprising the following families: CDK (cyclin-dependent kinase) [3], MAPK (mitogen-activated protein kinase), GSK (glycogen synthase kinase) [27] and CDK-like. There are six enzymes included in the eighteen *P. falciparum* CMGC kinases that are related more to CDKs than to other families [3].

Transmission of malaria via female mosquitos is another challenge faced in malaria control and eradication. The sexual stage of malaria is responsible for transmission of malaria to the mosquito when a female *Anopheles* mosquito bites an infected human [28]. CDPK4 is expressed in the gametocyte stage of the malaria parasite [29]. In the mosquito mid-gut, CDPK4 is an important signalling molecule for the transition of gametocyte into microgamete and macrogamete where the flagellated microgamete fuses with the macrogamete (exflagellation) to form a zygote [30,31]. *Plasmodium* CDPK4 is also important in maintaining the malaria life cycle and transmission [11,32]. Genomic information explained by Billker and his associates indicate that *Plasmodium berghei* CDPK4 (*Pb*CDPK4) is homologous to *Pf*CDPK4 (sharing 91% amino acid identity and 97% similarity) and *Toxoplasma gondii* CDPK1 (*Tg*CDPK1) sharing 73% identity and 88% similarity [31]. In other studies, it was demonstrated that *Pb*CDPK4 is essential for microgamete exflagellation and sexual stage development in the mosquito [2-4]. The gene disruption of CDPK4 in *P. berghei* caused severe defects in sexual reproduction of the parasite [31]. These studies have shown how important CDPK4 is in the biology of *Plasmodium* parasites. The currently available drugs, such as primaquine and artemisinin combination therapy (ACT) are also failing to stop malaria transmission from human to mosquito. The shortcoming of the two drugs is that primaquine is able to kill mature gametocytes but fails to kill immature gametocytes, while the opposite is the case for ACT [28]. It is important to discover compounds that will be able to inhibit CDPK4,

which is necessary for exflagellation in *Plasmodium* parasites. In a recent development, a chemical genetic approach validated *Pf*CDPK4 as a potential target of certain drug-like compounds blocking *Pf*CDPK4, inhibiting exflagellation and transmission [32].

1.1.4 Antimalarial natural compounds

Plants have been used as a source of medicines throughout history and continue to be used, serving as the basis for many pharmaceuticals used today. Natural products have played a significant role in the discovery of lead compounds for the treatment of malaria, leading to new drugs. Examples of various natural compounds that contributed to the development of effective antimalarial drugs include quinine [33], atovaquone, artemisinin (and its semi-synthetic derivatives) as well as clindamycin, erythromycin, azithromycin, chlortetracycline, tetracycline, oxytetracycline and doxycycline [34]. In recent years, there have been no antimalarial compounds originating from natural products. Many natural products have shown potent anti-*Plasmodium* effects but because of several medicinal chemistry reasons, have not been pushed forward into hit to lead drug discovery projects [35].

Compared to synthetic compounds, natural compounds offer a wide biologically relevant chemical space [36]. A review of the literature done to find compounds with antimalarial activities from natural products with the hope of developing effective drugs, clearly demonstrated that natural products can be used in many different ways. For example, as-yet unknown enzymatic activities due to a variety of biosynthetic pathways may be discovered, *in silico* tools can be used in natural product-based drug discovery [37] and libraries can be used for virtual screening [38]. Natural compound databases have been used in some virtual screening studies, revealing several active natural products as lead compounds [39].

1.1.5 *In silico* screening and its merits

Laboratory-based experiments for the study of compounds that target specific proteins for use as therapeutics can be costly and tedious. For most individual academic researchers and even research consortia, it would not be feasible to screen millions of molecules in the laboratory for binding to a given protein target or activity against a given pathogen. This is particularly true for researchers in the developing world where high-throughput screening facilities and skills are rare.

However, a computerised screening process with the aid of high-speed computers and some innovative software, can aid researchers in choosing and prioritising a smaller or reduced set of molecules for further drug development. Knowledge of molecular modelling, molecular biology and combinatorial chemistry is coded into software programs that help predict how the protein target and the drug molecule will interact [40]. Rational drug design and structure-based drug design (SBDD) are among several approaches used for drug discovery in combination with organic synthesis.

Rational drug design is a process used in the biopharmaceutical industry to discover and develop new drug compounds. It uses a variety of computational methods to identify novel compounds, design compounds for selectivity, efficacy and safety and develop compounds for clinical trial [40]. Rational drug design can be divided into two categories. The first one focuses on the development of small molecules with properties suitable for targets whose functional roles in cellular processes and structural information are known. This approach is applied extensively in the pharmaceutical industry. The second category aims at developing small molecules with predefined properties for targets whose cellular functions and their structural information may be known or unknown [41].

Virtual screening [42] is a computer-based method to search for novel compounds promising to be active against a chosen target [43] structure. The method aims at reducing a large number of hypothetical databases of chemical structures to a manageable set of compounds that can be screened *in vitro* as potential drug candidates [44]. Virtual screening methods have the advantage of reducing large libraries of chemical compounds to prioritized compounds, they are relatively inexpensive and reduce time in drug discovery processes [45]. There are two main approaches for virtual screening: Structure-Based Virtual Screening/Docking (SBVS) and Ligand-Based Virtual Screening/Docking (LBVS) [46].

Structure-Based Virtual Screening relies on the availability of the three dimensional structure of the target protein to predict receptor-ligand binding through docking of ligand databases against the active site of the target. The compounds in the databases screened are ranked to predict the binding affinity to select the most promising poses and therefore candidates from the database [47]. The 3D structure of a protein target bound to its natural ligand or a drug is most often derived from X-ray crystallography or nuclear magnetic resonance (NMR) techniques. This data in combination with mutagenesis and biochemical data can be used to identify its binding site, or the so-called active site. If the structure of the target is known, a library of chemical

compounds can be virtually screened and this enables identification of a potential new drug [41].

During Structure-Based Virtual Screening, docking and scoring are employed to screen large databases of compounds against the receptor to predict the ligand-receptor complex structure and binding affinity [44]. The docking process can be achieved by creation of ligand conformations within the protein active site followed by ranking conformations using various scoring functions [46] such as LigScore, PLP, LUDI and PMF [45].

The Ligand-Based Virtual Screening approach uses known active ligands to identify potential active compounds based on similarity search [48]. Unlike Structure-Based Virtual Screening, LBVS does not rely on the three-dimensional structure of the protein target and is therefore used only when the 3D structure of the target is not available [49]. The following approaches are also included in Ligand-Based Virtual Screening: similarity and substructure searching, quantitative structure-activity relationships [50], and pharmacophore and three-dimensional shape matching [47].

1.2 Aims and objectives

Emergence of antimalarial drug resistance has been reported for most of the current drugs used for treating malaria. Despite resistance to antimalarial drugs, drugs like chloroquine and sulphadoxine-pyrimethamine have been used extensively for treatment and prophylaxis of malaria [51]. In 2006, artemisinin combination therapy (ACT) was recommended by the WHO as first-line treatment of uncomplicated *falciparum* malaria in all areas in which malaria is endemic [52], but resistance has also been reported in combination therapy for treating *Plasmodium falciparum* malaria. This shows that current antimalarial drugs are not efficacious for malaria treatment. There is, therefore, an urgent need to develop novel drug targets and antimalarial drugs that are effective against drug-resistant malarial parasites.

An analysis by Ranjan and his associates confirmed that *Pf*CDPK4 is expressed in the *Plasmodium* gamete/gametocyte stage [24]. Disruption of the CDPK4 gene leads to severe defects in sexual reproduction in the parasite [30]. In addition the protein, or its homologues are completely absent in the human host. *Pf*CDPK4 is therefore a promising target for gametocyte-stage treatment of malaria. The structure of *Pf*CDPK4, in particular the kinase domain, can be used to discover new antimalarial drugs and to design chemical compounds or inhibitors that

may show anti-parasitic activity against the target enzyme. This study will focus on screening a large library of chemical compounds, including some natural compounds against the target enzyme, *Pf*CDPK4. In the future, this work could lead to the discovery of new antimalarial drugs that target the gametocyte stage.

The main aim of this study is to use a homology model structure of *Pf*CDPK4 for *in silico* screening of a large chemical compound library to search for anti-malarial lead compounds, and to validate *Pf*CDPK4 as a target for malaria drug discovery. The following objectives will be pursued:

(a) To build and validate a homology structure model of *Pf*CDPK4 using currently available experimentally determined homologue structures as templates.
(b) To use the *Pf*CDPK4 model structure as a target to virtually screen a large commercial library and a smaller library of chemical compounds from natural and synthetic sources.
(c) To use the *Pf*CDPK4 model structure as a target for virtually screening a small set of known clinically approved drugs that target protein kinases.
(d) To use information obtained from *in silico* virtual screening to prioritise a set of compounds for *in vitro* screening against cultured *P. falciparum*.

In the next chapters of this study, work aimed at fulfilling these objectives will be described. Homology modelling, Structure-Based Virtual Screening and scoring protocols will be described. Docking of libraries of chemical compounds against the *Pf*CDPK4 receptor and prediction of the binding affinity by scoring the resulting poses or conformations will also be discussed in detail. During the docking process, ligand poses within the protein active site were created and ranked using scoring functions. These methods were used to expedite the discovery of new hits and leads that may help to alleviate the current need for effective antimalarial drugs. Finally, the *in vitro* screening of several prioritized compounds for activity against the *Plasmodium* parasite late gametocyte stage will be described.

CHAPTER 2

MOLECULAR MODELLING OF *Pf*CDPK4

In silico drug screening and drug design involves the use of many tools that aid in making decisions at different steps during the drug discovery process, such as identification of a protein target of therapeutic interest, selection of new lead compounds and modification of them to obtain better affinities [53,54]. Before computational docking or *in silico* screening can be performed, a 3-dimensional molecular structure of the target protein must be obtained as a prerequisite for the structure-based docking procedure. Protein structures are commonly determined by experimental techniques such as X-ray crystallography or NMR spectroscopy. Not all desired target protein structures are readily available from the Protein Data Bank or indeed amenable to currently available experimental structure solution techniques. Sometimes a protein will not crystallise or is too large for NMR-based structure solution. The non-availability of the structure of the target protein can be solved by implementation of computational methods like threading and homology modelling which, if done correctly, can be employed to reliably predict the structure of the desired protein [50].

The homology modelling method is commonly used to determine or predict the 3D structure of a protein based on template proteins with closely homologous structures [55,56]. Homology modelling involves several successive steps including template identification, sequence alignment between the target and template proteins, model building, model refinement and validation of the model [50]. The Basic Local Alignment Search Tool (BLAST) search [57] is one of the tools used to search for homologue(s) of known protein structures that are similar to the target sequence to be modelled. BLAST and PSI-BLAST are profile-based programs that use a repetition procedure to search for proteins with distant similarity [58] to the query sequence (target sequence) [57]. The best available template structure can be selected based on a number of criteria including percentage identity (or similarity) to the target, resolution of structure, and availability of ligand-bound structures. Sometimes it is prudent to use several structures as template. When the best template or template structures have been identified, the sequence of the template(s) can be aligned with the sequence of the target to be modelled. With the aligned sequences readily available, models can be built using various different tools such as MODELLER [59] (http://salilab.org/modeller/), WHATIF (http://swift.cmbi.ru.nl/whatif/), SWISS-MODEL [60], and Protein Model Portal [61].

Different modelling programs use different protocols to build a three-dimensional structure of a known target [62]. MODELLER is a computer program for comparative building of homology models. MODELLER uses a query or template structure and the alignment between the sequence of the desired target with that of the template structure to build a homology model structure of the target [59,63]. The model structure created often needs to be refined by energy minimization or other molecular dynamics simulations using tools such as CHARMm [64], AMBER [65], or GROMOS [66], which can provide a mechanics and dynamics protocol for studying energetics and motions of molecules [67,68].

2.1 Materials and methods

2.1.1 Structural prediction of *Pf*CDPK4 through homology modelling

The amino acid sequence of the kinase or ATP-binding domain of *Pf*CDPK4 was retrieved from UniProt Knowledgebase (UniProtKB) with the Uniprot ID Q8IBS5 [69]. Since *Pf*CDPK4 does not have a solved 3D structure, the amino acid sequence of *Pf*CDPK4 was used as a query sequence to search for a template structure from the Protein Data Bank (PDB) through BLAST Search [57] via Discovery Studio 3.5 (Accelrys) (Appendix A). The BLAST Search protocol searches within the PDB_nr95 sequence database in Discovery Studio, with BLOSUM62 chosen as substitution/comparison (scoring) matrix. The search method identifies the best template by initially aligning the sequence of the target (*Pf*CDPK4) with the sequences of proteins with known structures. A template with a sequence that has close sequence identity to that of *Pf*CDPK4 was selected from the hit list. This template was the kinase or ATP-binding domain *Tg*CDPK1 (PDB Id: 3ma6) from the single-celled parasite *Toxoplasma gondii*. Using 3ma6 PDB Id, the structure of *Tg*CDPK1 was obtained from the RCSB Protein Data Bank [70] (http://www.rcsb.org/pdb/).

The amino acid sequence of *Pf*CDPK4 was aligned with the sequence of *Tg*CDPK1 using the Align 123 tool in Discovery Studio 3.5. An alignment gap penalty of ten was used in the progressive pairwise alignment algorithm modified from the CLUSTAL W [71] program. The alignment obtained between *Pf*CDPK4 and *Tg*CDPK1 sequences was used to build homology models using the MODELLER 9v8 [72] engine operated within Discovery Studio 3.5. From the Build Homology Model protocol, a total output of twenty models was selected. The list of 20 models produced was evaluated using the Verify Protein (MODELLER) protocol and the

Verify Protein (Profiles-3D) protocol to help select the best quality model. The Verify Protein (MODELLER) protocol allows for the selection of the best structure from a collection of protein structures with the same amino acid sequence, and the DOPE (Discrete Optimized Protein Energy) Score calculated for each protein structure is used to compare different conformations of a protein. Statistically, a lower DOPE Score indicates a better model [73]. DOPE measures the quality of the protein structure based on an atomic distance-dependent statistical potential calculated from a sample of native protein structures [74]. The Verify Protein (Profiles-3D) protocol evaluates the compatibility of the three-dimensional structure of a protein model with the sequence of residues it contains. The Verify Score measures the overall quality of the model structure [73]. The model protein structure with the lowest possible DOPE Score and highest possible Verify Score was selected for further modification and validation.

2.1.2 *Pf*CDPK4 model structure refinement and validation for stereochemical quality

The selected model (*Pf*CDPK4.F06) structure was refined to search for low energy conformations of loop regions using the Loop Refinement tool, which uses CHARMm based molecular mechanics. Protein loops are flexible regions on protein surface [75] and are parts of the polypeptide that do not have secondary structure [76]. For each loop selected, several low energy conformations were generated and the best loop refined model structure was selected after assessing the DOPE Score and Verify Score of each conformation. Residues with weak similarities or no similarities for the alignment between *Pf*CDPK4.F06 model and *Tg*CDPK4 template during model building were refined. The refined models were analysed using Verify Protein (MODELER) and Verify Protein (Profiles-3D). In total, the following amino acids were loop refined: Glu132-Ile140, Gly196-Asn204, Leu55-Lys62, Met54-Lys62, Asn117-Lys125, Lys33-Glu43, Leu55-Lys62, Gly215-Leu221, His151-Lys157, Val119-Pro126, Met240-Ser248, Phe152-Lys157, Asp78-Gly82, Lys33-Asp41, Ile144-Gly147, Asp19-His24, Lys225-Ser228, Gly9-Phe11 and Asp69-Asn71. The final model was validated using PROCHECK [60, 77] through the SWISS-MODEL Workspace website (http://swissmodel.expasy.org/workspace/). The stereochemical quality of the final model (*Pf*CDPK4) was analysed from the Ramachandran Plot produced. This plot was introduced by Ramachandran in 1963 [78] as an approach to check the geometric quality of protein structure, and has been widely used since. The Ramachandran angles of the polypeptide chain describe

the rotation of amino acids and proteins around the bonds between the C-N-Cα–C (Phi, φ) and N-Cα-C-N (Psi, ψ) torsion angles of the backbone conformation [78,79]. The three-dimensional structure of a protein model is validated by the Ramachandran plot by checking the map for residues that fall into allowed regions, generously allowed regions, low-energy regions and disallowed regions [77]. A protein model structure with most of its amino acid residues in the allowed regions indicates that the model structure is good. A poor protein model structure will have many of its amino acid residues in the disallowed regions [79]. The WHAT-IF online server was used to determine the quality of the geometry by calculating the Ramachandran Z-Score (ProCheck) (http://swift.cmbi.ru.nl/servers/html/index.html).

2.2 Results and discussion

2.2.1 Structural prediction of *Pf*CDPK4 through homology modelling

The complete sequence of *Pf*CDPK4 retrieved from the UniProt Knowledgebase (UniProtKB) [69] consists of five domains with a sequence length of 528 amino acid residues. The sequence of the domain of interest, protein kinase, was chosen from the *Pf*CDPK4 sequence position 71 to position 329, resulting in a domain 259 residues in length.

A BLAST search with this sequence found a list of 240 homologues of the *Pf*CDPK4 kinase domain with known protein structures in the Protein Data Bank. From the hit list, five templates (Table 1) were found to have suitably good sequence identity and similarity to the *Pf*CDPK4 sequence. The level of similarity between the template and target sequences is important as it is strongly linked to model quality [80]. The least expected sequence identity between the template and the target is 30% for preparation of a reliable model [81,82]. Therefore, building a model with template-target sequence similarity over 50% suggests that the model will be accurate enough for the further drug discovery process [56,80]. Calmodulin-domain protein kinase 1 (PDB Id: 3ma6) from *Toxoplasma gondii* (*Tg*CDPK1) was selected as a template because of its close sequence identity to that of *Pf*CDPK4 with 79% sequence identity. Besides having good sequence identity, template 3ma6 also has a good resolution of 2.5 Å. Its X-ray crystal structure was downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/).

Table 1: Templates with the highest percentage sequence identity from a list of 240 BLAST Search hits.

PDB id	Sequence Identity (%)	Sequence Length	Alignment Length	Bit Score	E-Value	Resolution (Å)	Organism
3ma6 (B)	79	266	259	427.56	1.02e-120	2.50	*T. gondii*
3nyv (A)	78	457	259	426.79	1.85e-120	1.88	*T. gondii*
2wei (A)	68	278	257	362.07	6.61e-101	1.65	*C. parvum Iowa II*
3lij (A)	63	465	257	323.17	3.03e-89	1.90	*C. parvum Iowa II*
3dxn (A)	61	258	250	315.08	9.35e-87	2.17	*T. gondii*

The final alignment between the *Pf*CDPK4 and *Tg*CDPK1 amino acid sequences showed 78% sequence identity and 90% sequence similarity (Figure 3).

Figure 3: Alignment between *Pf*CDPK4 and TgCDPK1 showing 78% sequence identity and 90% sequence similarity. Dark-Blue colour shows conserved amino acid residues, Light-Blue – strong similarity, Faint-Blue – weak similarity, White – no similarity.

The alignment was sufficiently good for use as a basis for building the 3D structure model of *Pf*CDPK4 using MODELER engine in Discovery Studio 3.5. A maximum of 20 models of *Pf*CDPK4 were generated and the best model of *Pf*CDPK4 was selected based on the lowest DOPE Score and the highest Verify Score (Table 2). Two insertions (Arg38 and Glu137) were found in the models generated; these are residue groups of the model that did not align with the template residues. The following amino acid residues were found not to be similar to template residues and are amongst the loop-refined amino acids: Ile3, His21, Tyr26, Met54, Ile58, Tyr96, Met139, Ile140, Tyr154, Cys186, Ala214 and Met240.

*Pf*CDPK4.F06 model was selected as its DOPE Score was lowest (-30 087) among the model structures listed in Table 2. The Verify Score of model *Pf*CDPK4.F06 was found to be the highest (113) making it the best model from all the models created. However, for the model structure to be of better quality, its Verify Score should be higher than the expected high score (117.717) for proteins of similar size. None of the models generated had a Verify Score higher than the expected high score leading to a decision to refine the model to improve its quality. Since the Verify Score is the sum of the scores of all residues in the model protein, the Verify Score of each amino acid residue indicates whether a residue is in the desired 3D environment.

Table 2: Models generated by MODELLER and their scores (DOPE Score and Verify Score) as a measure for the quality of the models.

Name	DOPE Score	Verify Expected Low Score	Verify Expected High Score	Verify Score
*Pf*CDPK4.F01	-29 966	52.973	117.717	112
*Pf*CDPK4.F02	-29 834	52.973	117.717	108
*Pf*CDPK4.F03	-29 830	52.973	117.717	106
*Pf*CDPK4.F04	-29 576	52.973	117.717	110
*Pf*CDPK4.F05	-29 639	52.973	117.717	108
*Pf*CDPK4.F06	-30 087	52.973	117.717	113
*Pf*CDPK4.F07	-29 718	52.973	117.717	106
*Pf*CDPK4.F08	-29 819	52.973	117.717	112
*Pf*CDPK4.F09	-29 944	52.973	117.717	114
*Pf*CDPK4.F10	-29 937	52.973	117.717	110
*Pf*CDPK4.F11	-29 792	52.973	117.717	108
*Pf*CDPK4.F12	-29 924	52.973	117.717	108
*Pf*CDPK4.F13	-29 416	52.973	117.717	103
*Pf*CDPK4.F14	-29 651	52.973	117.717	107
*Pf*CDPK4.F15	-29 933	52.973	117.717	112
*Pf*CDPK4.F16	-29 719	52.973	117.717	107
*Pf*CDPK4.F17	-29 612	52.973	117.717	103
*Pf*CDPK4.F18	-29 767	52.973	117.717	110
*Pf*CDPK4.F19	-29 678	52.973	117.717	106
*Pf*CDPK4.F20	-29 786	52.973	117.717	112

2.2.2 *Pf*CDPK4 model structure refinement and validation for stereochemical quality

The Profile-3D line plot (Figure 4) for *Pf*CDPK4 amino acid residues shows that Leu130 was in the negative region and Leu131, Asn135 and Val245 were closely approaching zero. This plot analyses amino acid residues for their compatibility in favouring three-dimensional structure. Irregular folding of amino acid residues is mostly responsible for the decrease in Verify Scores, while amino acid residues with a Verify Score above zero show stable folding and hence favour the three-dimensional structure [83]. As a result of the observed low scoring amino acids, loop refinement of selected residues was performed.

After optimization of the *Pf*CDPK4.F06 model, the DOPE Score and the Verify Score of the refined model improved (Table 3). The Verify Score of the refined model was 126, which is higher than the expected high score of 117.717. The amino acids with low Verify Scores were now in the reasonable score region (Figures 4), which implied an improvement of the final model quality.

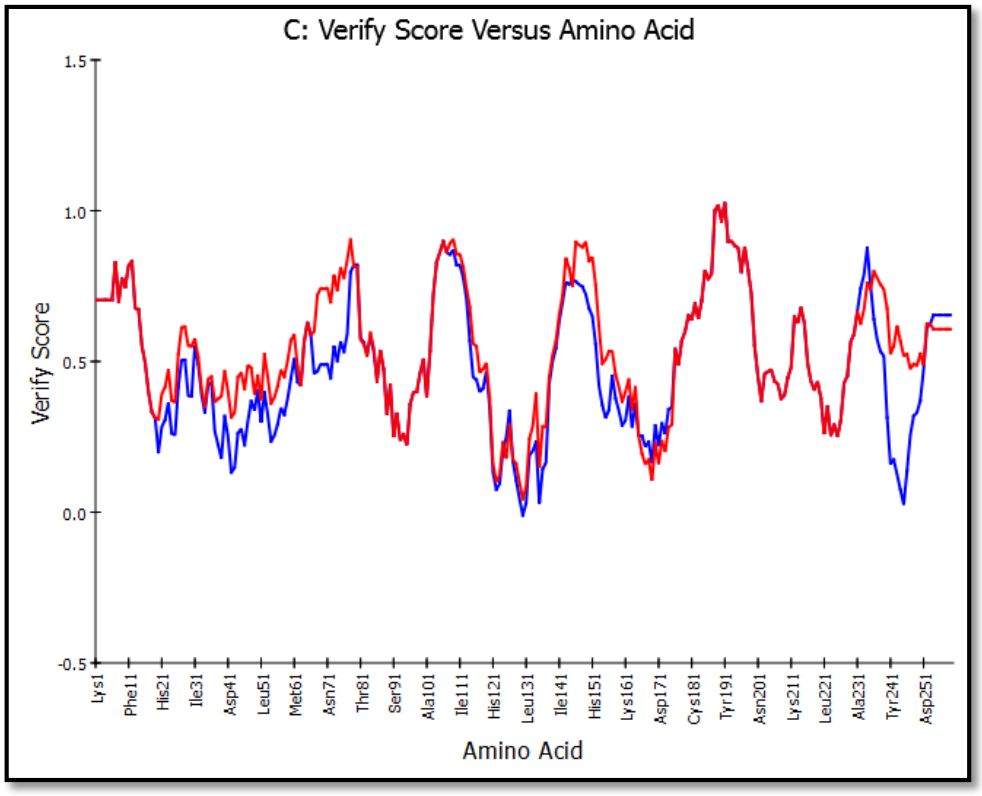


Figure 4: Verify Protein (Profiles-3D) line plot for initial model (*Pf*CDPK4.F06) in blue and final model (*Pf*CDPK4) in red. It is clear from the plot that the amino acid Verify Score for the final model is better than that of the initial model.

Table 3: Comparison of the quality and the availability of the initial model and the final refined model of *Pf*CDPK4. The Ramachandran Z-score determines the quality of the Ramachandran plot. The scores were within the expected ranges for well-refined structures.

Name	Dope Score	Verify Score	Ramachandran Z-Score
Initial Model (*Pf*CDPK4.F06)	-30 087	113	0.800
Final Model (*Pf*CDPK4)	-30 813	126	0.072

The final model structure of *Pf*CDPK4 was superimposed on the template structure (*Tg*CDPK1, PDB id: 3ma6, chain B), showing a very close similarity between the two structures with a Root Mean Square Deviation (RMSD) of 0.92 Å (Figure 5).

Figure 5: The superimposition of the homology model structure of *Pf*CDPK4 (purple) and the template structure, *Tg*CDPK1 (blue).

Final model validation was done using PROCHECK which determines the stereochemical quality of the model through Ramachandran plots. The Ramachandran plot analyses and displays the backbone conformational angles for all the amino acids in the protein structures. The Ramachandran plot produced by PROCHECK [60] through the SWISS-MODEL

Workspace online (http://swissmodel.expasy.org/workspace/) (Figure 6) indicated that 96% of the phi and psi angles of amino acids in the created final model (*Pf*CDPK4) lie within the most favoured or low-energy regions (A, B, L), 3.4% within the additional allowed regions (a, b, l, p), 0.4% in generously allowed regions (~a, ~b, ~l, ~p), while only 0.4% lie within disallowed regions (Figure 6). The quality of the geometry was further confirmed by calculation of the Ramachandran Z-Score (Table 3) through the WHAT-IF online server. The results from these analyses show that the model structure of *Pf*CDPK4 is of sound stereochemical and geometrical quality.

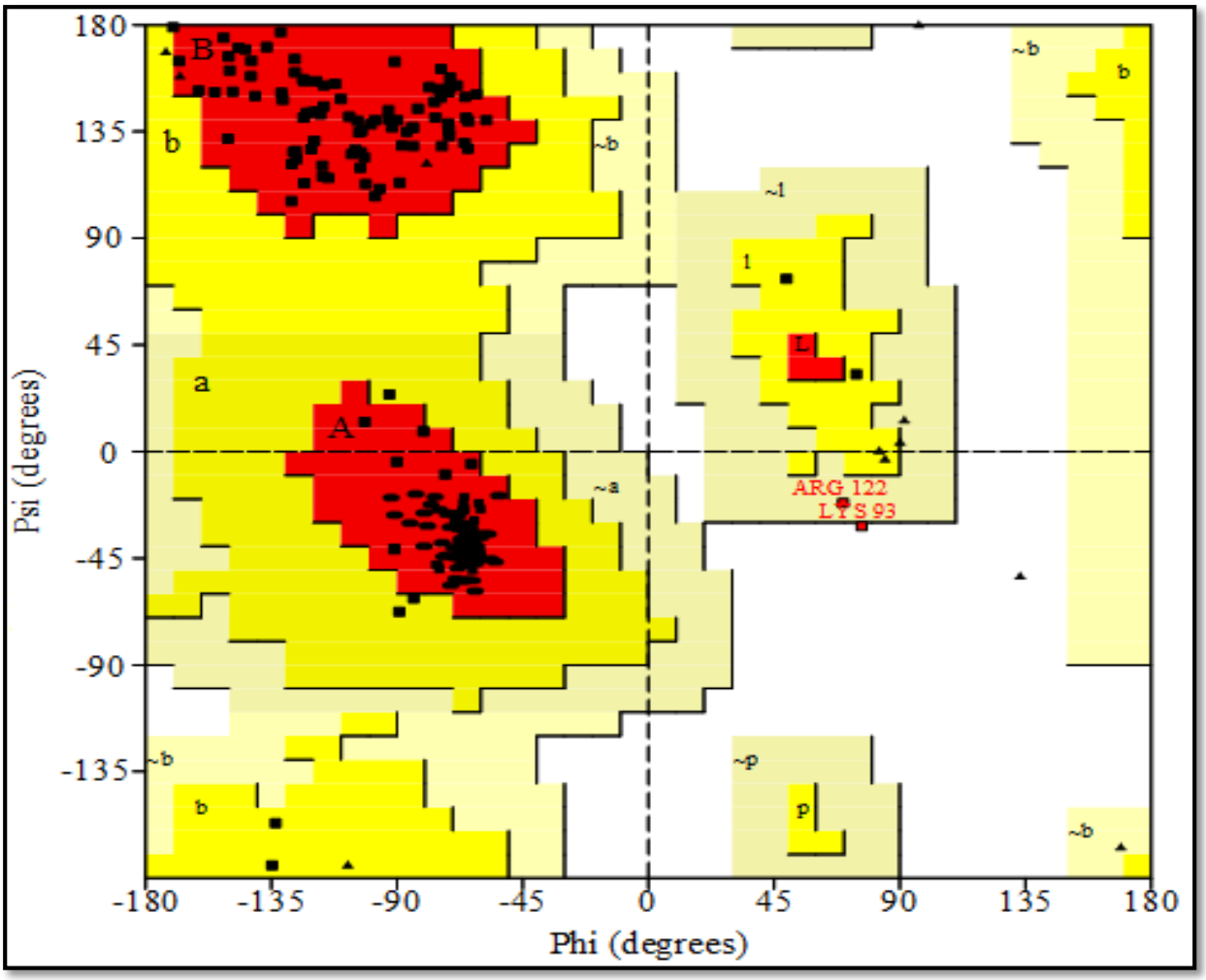


Figure 6: Ramachandran plot of the *Pf*CDPK4 model using PROCHECK. The regions for amino acid residues in the plot are represented by colours as follows: red – most favoured region, brown – additional allowed region, yellow – generously allowed region and pale-yellow – disallowed region.

CHAPTER 3

TARGET AND LIBRARY PREPARATIONS

Before docking or *in silico* screening, the target protein structure file and the ligand structure files need to be prepared. Preparing the target may involve adding missing hydrogen atoms, searching for or identifying a binding site and editing the binding site where necessary. A binding site is a small pocket/region where a ligand is most likely to bind to produce the desired effect including activation, inhibition or catalysis [84]. Identifying the correct target binding site is a crucial step in *in silico* screening and structure-based drug design approaches.

3.1 Introduction

It is often assumed that the binding sites of a protein are characterized by a large cavity that gives the protein some advantage, although this is not always the case. According to Laskowski *et al.* [85], there are three advantages for proteins having a large cavity: a large cavity provides a maximum number of hydrogen-bonding and hydrophobic interactions. The second one is that it enables precise positioning of the substrate, which facilitates catalysis. Thirdly, burying a substrate in a cavity provides shielding from polarized water molecules, hence decreasing the dielectric constant and allowing the protein to generate the strong electrostatic forces necessary for catalysis [85]. However, the binding sites of proteins have various shapes. For example, the binding site of endonuclease is a spherical cavity with deeply-buried ligand whereas the binding site of ribonuclease is an elongated groove containing exposed ligand. Large proteins tend to have many binding sites, but this does not mean that they have larger binding sites [86]. Several computational methods can be used to identify the binding site of the target. These methods can be roughly categorized into sequence-based, template-based, geometry-, and energy-based approaches [87,88]. The most commonly used methods are geometry-based and energy-based [86,89].

The energy-based approach predicts the pockets or cavities by computing the interaction energy between the protein atoms and a small-molecule probe group [86,90]. GRID [91] and Q-SiteFinder [92] are used together where Q-SiteFinder uses the GRID force field to calculate the hydrogen bond energy between the protein target and the ligand probe groups (like $-CH_3$), and cluster analysis is carried out to identify regions with the highest total interaction energy. The

highest total interaction energy mostly relates to the binding site [87,88]. The energy-based method is most suitable for a protein with a co-crystallized ligand, which provides the required information about the volume of a known ligand pose that has occupied the active site of the target structure [93]. From the generated grid points, those selected lie within a specified radius of the ligand atoms. The active site is determined as a collection of all grid points occupied by the ligand [94].

The geometry-based approach requires a knowledge of the 3D protein structure and its aim is to detect solvent accessible grid points that are embedded in the protein surface [86]. This method is further divided into three types, which are: grid system scanning, probe sphere filling, and alpha-shape. The POCKET [95] and LIGSITE [96] algorithm methods are commonly used in the geometry-based approach [87,90], and these algorithms do not require any knowledge of the ligands [88]. The geometry-based method uses an "eraser" algorithm that identifies the cavities within the target by taking into consideration the shape of the receptor. Grid points are constructed in a rectangular shape around the receptor. Each of these grid points can be categorized as occupied or free. Grid points that are occupied fall within the set contact distance of the nearest protein atom while free (unoccupied) grid points are outside the contact distance. The eraser algorithm is employed to remove grid points that are not within contact distance of the nearest protein atoms, resulting in sites being isolated from the remaining grid points. The remaining free grid points are then grouped as possible active sites [94] using a flood-filling algorithm.

Before ligand molecules can be used in docking, it is important to prepare them. Ligand libraries or ligand databases can be obtained from different sources, such as ZINC (http://zinc.docking.org), PubChem (http://pubchem.ncbi.nlm.nih.gov/), ChEMBL (http://www.ebi.ac.uk/chembl), and ChemBank (http://chembank.broadinstitute.org/). Compounds can be filtered for drug-likeness properties [27], reducing or eliminating compounds that are deemed not suitable for drug development [97]. Drug-likeness is defined as a compound having sufficiently acceptable ADME (absorption, distribution, metabolism, and elimination) properties and sufficiently acceptable toxicity properties to successfully complete human Phase I clinical trials [98]. Lipinski and his colleagues discovered that for a drug to be soluble and membrane permeable, it should pass the following criteria: the compound should have less than five hydrogen donors, its molecular weight should be less than five hundred, its log*P* less than five, and the number of hydrogen-acceptor groups less than ten [99]. In order to

meet the above-mentioned criteria, ligands need to be filtered based on these rules (the 'Lipinski rules') to compensate for drug-like properties.

3.2 Materials and methods

3.2.1 *Pf*CDPK4 model optimization and binding site identification

The model structure of *Pf*CDPK4 created as described in Chapter 2 above was prepared using the "prepare protein" protocol within Discovery Studio 3.5 (Accelrys), which cleans up problems in the input protein structure file by running steps like protonation, standardizing atom names, inserting missing atoms in residues and removing alternative conformations.

The "prepare protein" protocol was run to ensure that missing hydrogen atoms are inserted into incomplete amino acid residues by protonation at pH 7.4 while applying the CHARMm force field. The binding site was defined from the receptor's (*Pf*CDPK4) cavities. The method identified the binding site using an erase algorithm based on the shape of the *Pf*CDPK4 structure. The cavity detection method works through the following steps (Figure 7): The 3D rectangular grid is created with the default grid spacing of 0.5 Å. The grid points are scanned along the *x, y, and z* axes with the minimum number of grid points set to 100. The radius for receptor hydrogen atoms was set to 2.0 Å and the radius for receptor heavy atoms was set to 2.5 Å. Grid points that fall within the radius (2.0 Å) of all hydrogen atoms and within the radius (2.5 Å) of all heavy atoms in the receptor were marked as occupied. Grid points lying outside the radius (2.0 Å) of all hydrogen atoms and outside the radius (2.5 Å) of all heavy atoms of the receptor are marked unoccupied (free). A cubic shaped eraser with site opening set to 5 Å was employed starting from the grid, removing any grid points it comes in contact with without touching the receptor atoms. The site opening value corresponds to the maximum size of the mouth of the site. Flood-filling is used to group together the remaining grid points belonging to a given well-defined binding site [94].

The available tools were used to expand or contract both the binding site and its surrounding sphere. The binding site can be increased by adding free grid points adjacent to the binding site and can be decreased by removing grid points. The binding site within the *Pf*CDPK4 model structure was edited by contracting the surrounding sphere and the grid points, reducing the sphere diameter from 17.8 Å to 11 Å and removing the outermost site points that were adjacent to free grid points from the binding site. Protruding binding site points were manually removed by selecting the parts that protrude and then deleting them.

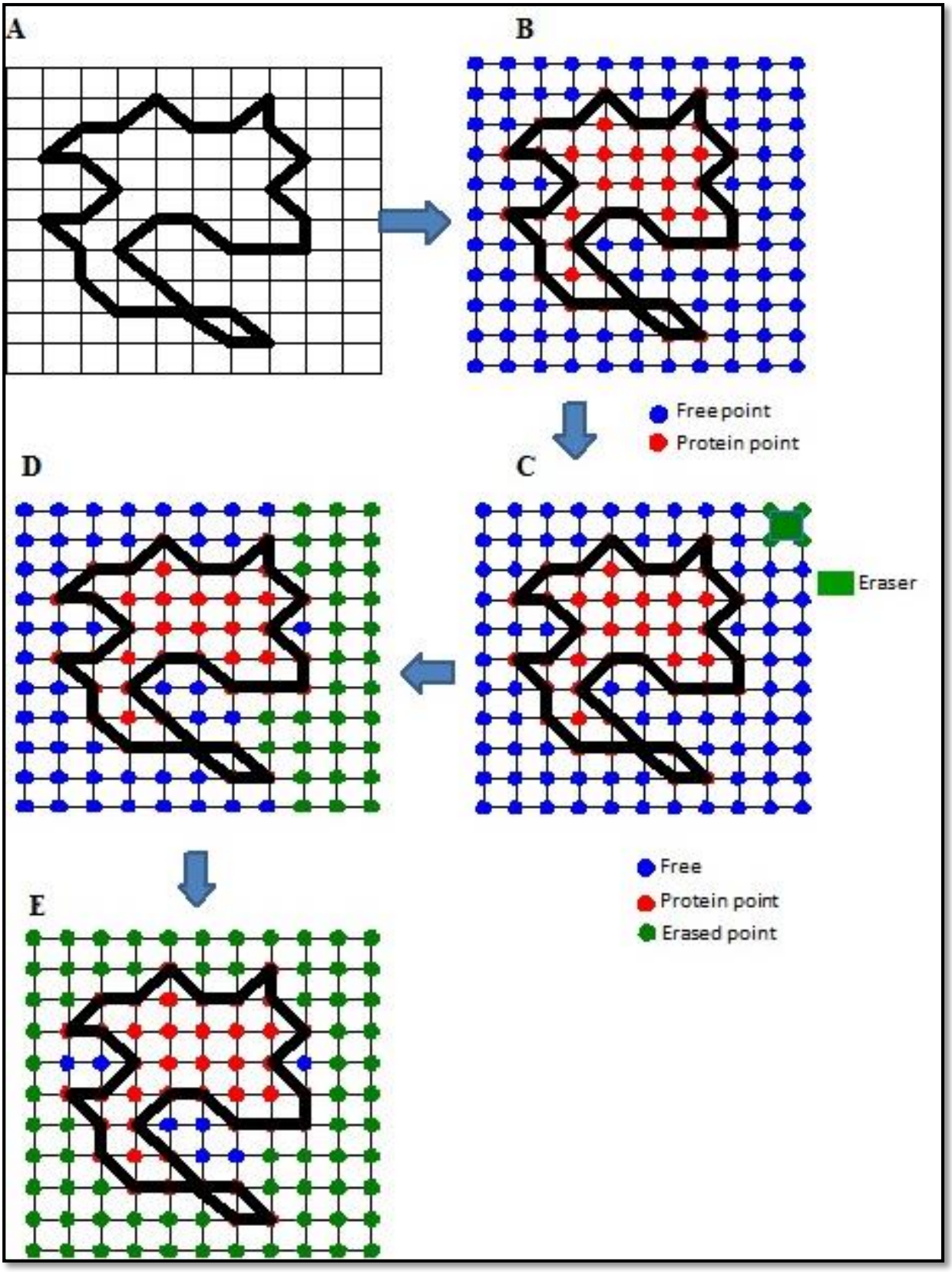


Figure 7: Representation of how eraser algorithm operates for detecting binding site from protein surface. A: Grid creation (0.5 Å grid spacing), B: Probe used to test for Van der Waal clashes at each grid point, C: Eraser removes free points and D: All points untouched by the eraser are considered as a site.

3.2.2 Preparation and optimization of Bio-Focus library, NSC library and kinase inhibitors

Three sets of libraries were available for the current study. These were the Bio-Focus library consisting of 20 000 compounds, a kinase inhibitor library, which is a small subset of 13 clinically approved compounds and the NSC (Natural and Synthetic Compounds) library, which is a collection of 28 natural and synthetic compounds known or expected to exhibit antimalarial activity. Compounds from the Bio-Focus and NSC libraries were filtered using the Lipinski Rules in Discovery Studio tools. The library of 13 clinically approved compounds was not filtered because the compounds had already been approved by the FDA (Food and Drug Administration). Compounds that were accepted by the Lipinski Rules were prepared using the Prepare Ligands protocol, which optimizes compounds for docking. During ligand preparation, compound tautomers and isomers were generated, bad valencies were fixed and duplicates were removed. The clinically approved kinase inhibitors were also optimized using the Prepare Ligands protocol.

3.3 Results and Discussion

3.3.1 *Pf*CDPK4 model optimization and Binding site identification

During model validation, no major errors were found in the *Pf*CDPK4 model. The Prepare Protein protocol improved the structure prior to *in silico* docking by adding hydrogens to optimize the protonation state of individual amino acids. The ATP-binding site of the model structure was identified by searching the cavities within the target model using "eraser" algorithms, as explained above. The search found seven binding sites of different sizes and shapes. The binding sites were displayed as a set of grid points and a transparent sphere (Figure 8).

Protein kinases are responsible for the transfer of phosphate groups from adenosine triphosphate (ATP) to the hydroxyl group of serine, threonine or tyrosine in the substrate [100]. From seven identified binding sites, only one was selected based on knowledge from other protein kinase ATP binding sites (Figure 9). The ATP binding site is similar in different kinases, and most kinase inhibitors are ATP competitive, making selectivity and promiscuity a potential problem [101].

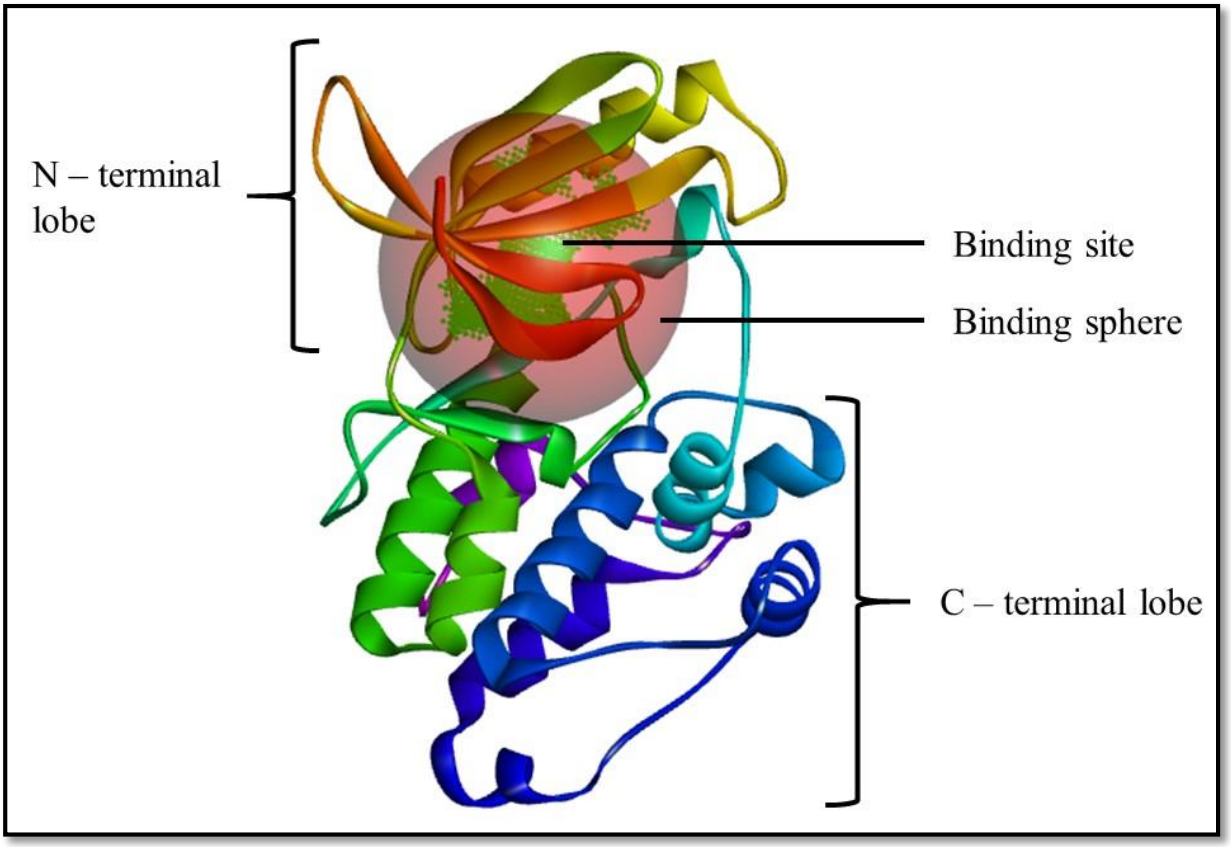


Figure 8: The final model structure of *Pf*CDPK4 with its binding site in green dots and its transparent surrounding sphere in light red.

The kinase catalytic domain has two lobes (N-terminal lobe and C-terminal lobe) joined by a segment called a hinge loop. The N-terminal lobe is small and consists mostly of five antiparallel β-sheets and a conserved α-helix while the C-terminal lobe is mainly α-helical. The cleft formed between the two lobes constitutes the active site region, which contains the residues that are directly involved in either catalysis or ATP binding [102]. Vulpetti and Bosotti identified a set of 38 residues within the five ATP binding pockets that make up the ATP binding site of protein kinases [103], while Huang *et al.* have shown that the five ATP binding pockets are lined with 36 residues [101]. The five ATP binding pockets are the adenosine binding region (kinase hinge region), hydrophobic pocket, phosphate binding pocket, ribose binding pocket and entrance pocket [101]. Based on this information, selecting the right binding site in the current study was unproblematic, and the selected binding site was the largest of all seven binding sites that were identified.

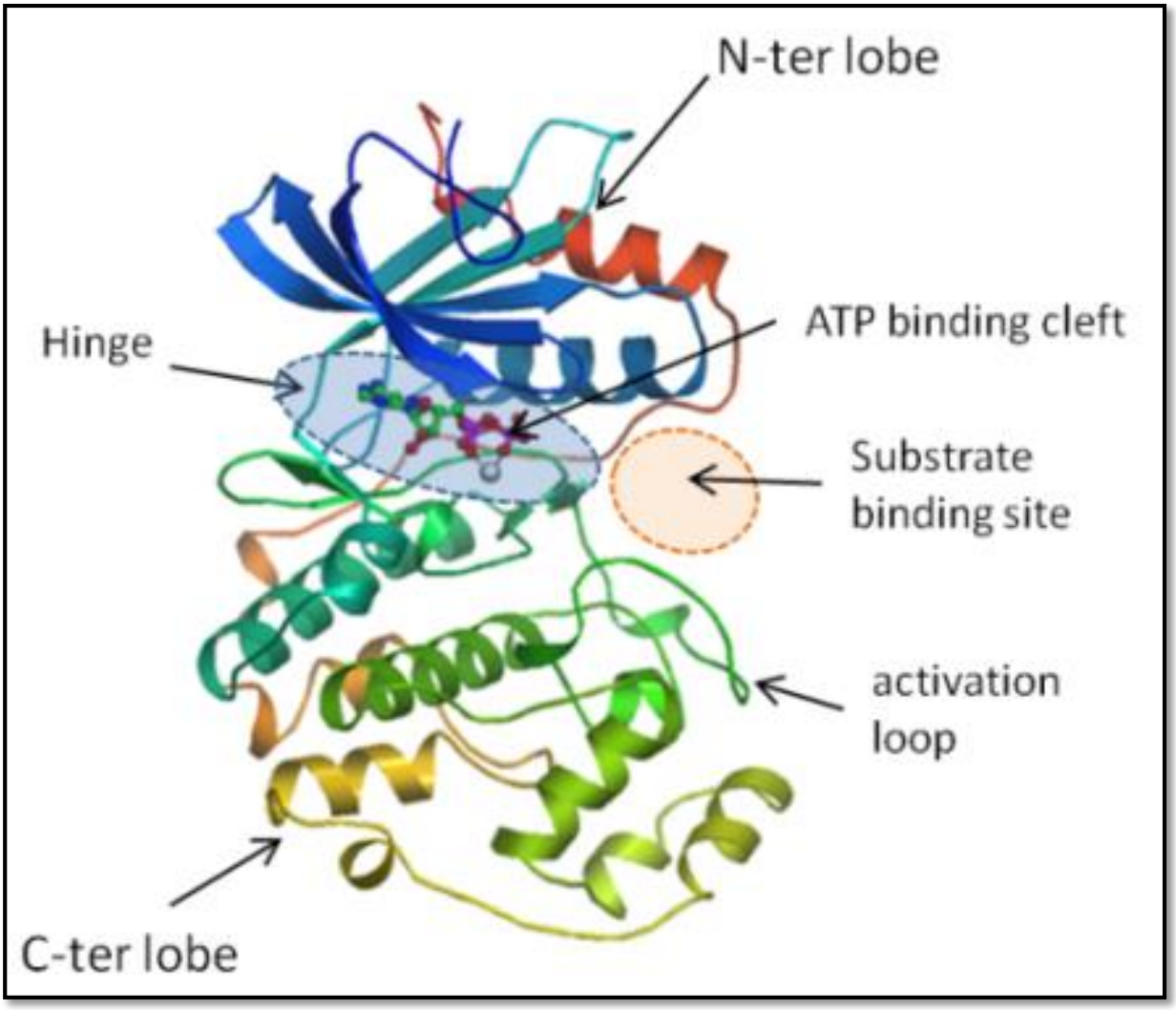


Figure 9: A schematic depiction of the ATP binding site location in one of the known kinase protein structures, ERK2 (Extracellular Signal-Regulated Kinase 2) [2].

3.3.2 Preparation and optimization of Bio-Focus library, NSC library and kinase inhibitors

The Lipinski Rule of Five passed 19 351 compounds from the Bio-Focus library and failed 649. From the NSC library, 23 compounds passed the Lipinski Rule of Five and five compounds failed. Preparation of the 19 351 Bio-Focus compounds resulted in 83 707 compounds while preparation of the 23 NSC compounds resulted in 37 optimized compounds. Preparation of the 13 clinically approved kinase inhibitors resulted in 36 optimized compounds. The increased number of prepared compounds in all the libraries compared to the original number of compounds was due to the generation of 3-dimensional structures of tautomers, stereoisomers and ionized forms.

CHAPTER 4

IN SILICO VIRTUAL SCREENING AND *IN VITRO* SCREENING OF SMALL MOLECULES AGAINST THE PUTATIVE DRUG TARGET *Pf*CDPK4

High-throughput screening has been used previously and continues to be used in the early stages of drug design. It is a process of screening large physical libraries of compounds in order to find potentially active compounds against a specific target [53]. Screening large physical libraries in a small space of time and with reduced effort is the main purpose of High-Throughput Screening [104]. Though High-Throughput Screening (HTS) has been used for many years as a standard method for drug discovery in the pharmaceutical industry, there are several challenges and disadvantages encountered in HTS. Some of the challenges encountered with HTS is that it is time consuming, expensive and has to date failed to find leads for many targets [105].

4.1 Introduction

Virtual High-Throughput Screening or virtual screening is a computational method that identifies ligands with a specific binding affinity for the target binding site from a large library of chemical compounds. Virtual High-Throughput Screening, or molecular docking, can therefore be used to test possible hypotheses before conducting costly and risky laboratory experiments. There are two main types of virtual screening, classified as structure-based virtual screening and ligand-based virtual screening [84]. Compared to laboratory-based screening experiments, the advantage of virtual screening is that it is carried out at relatively lower cost since initially no compounds have to be purchased or synthesized.

Docking approaches are also used in SBDD, which relies on the knowledge and the availability of the 3D structure of the target protein and its binding site to investigate the interaction between the small-molecule ligand and the receptor [83]. The aim of molecular docking is to predict the position, orientation and conformation of a small-molecule ligand within the binding site of a target protein [106]. To be able to carry out docking there are several prerequisites that should be met. The prerequisites of docking include the receptor protein 3D structure, a library

of small-molecule ligands or chemical compounds, and a computer program that is able to perform docking and scoring procedures [45].

Many popular early docking mechanisms treated the ligand as flexible while the protein conformation was kept rigid; this relies on the 'lock-and-key' model of protein-ligand binding. It is now widely accepted that ligand binding is not a static event but a dynamic process, in which both the ligand and receptor may undergo conformational changes [107]. Thus, the induced-fit theory suggests that both the ligand and the receptor should be treated as flexible during docking, which would describe ligand-receptor binding more accurately [46]. Consequently, docking algorithms should handle the flexibility of both molecules (ligand and receptor). Many docking programs can apply some flexibility to the protein during the docking through active site side chain rotations [53].

There are many computer programs or methodologies capable of docking small molecule ligands into the receptor's active site. Such docking programs or methods include GOLD [108], AutoDock [109], FlexX [110] and others. For the present project, two docking methods were considered: LibDock [111] and CDOCKER [112]. LibDock is one of the available methods that are able to rapidly predict the ligand pose using the molecular docking algorithm. The LibDock algorithm is capable of screening a large library of ligands against the receptor within reasonable computation times. The LibDock algorithm work flow starts by searching conformations of ligands, generating hot spots, matching the hot spots and the ligand molecule, and finally optimization of the complex followed by scoring [111].

CDOCKER [112,113] is a grid-based molecular docking method that employs CHARMm to generate random ligand conformations through high temperature molecular dynamics. The advantage of CDOCKER is that it allows full ligand flexibility, application of force fields, reasonable computation time and the introduction of soft-core potentials which reduce bad contacts [112]. Screening using CDOCKER allows the ligand to be fully flexible while the target is held rigid. The method generates random ligand conformations through high-temperature molecular dynamics during the initial placing of the ligand into the active site [83]. The poses are searched using random (Rigid-Body) rotations, which are followed by grid-based simulated annealing. The simulated annealing algorithm consists of a set of multiple heating steps at high temperature followed by multiple cooling steps at low temperature to find minimal energy conformations. This is followed by full minimization and output of refined ligand poses.

This method is suitable for screening small ligand libraries and for refining docking poses [112] due to the relatively expensive computational processing requirement.

Ranking and scoring of the docked conformation is very important in structure-based virtual screening to differentiate true positives from false positives [114]. Scoring aims at identifying the correct binding pose of protein-ligand complexes according to their binding affinities. However, scoring is one of the major challenges in molecular docking and much has been done in the development of accurate scoring functions [45]. Docking often includes the effective search of conformational space through a posing mechanism where the ligand is placed into the target protein in different orientations to be able to identify the correct binding mode of the ligand molecules [84]. Several scoring functions are typically used to account for the distinction of true binding modes as it is not reliable to use a single scoring function to rank every protein-ligand complex. Scoring functions can be grouped into three main types: force field-based scoring functions, empirical scoring functions, and knowledge-based scoring functions [115].

Force field-based scoring functions account for the binding energy between ligand and protein by evaluation of non-bonded interactions. The energies are usually estimated through a combination of van der Waals potentials with an electrostatic energy term [116].

Empirical scoring functions are based on a set of uncorrelated terms characterizing various aspects of protein–ligand interactions; these terms include energies of hydrogen bonding, steric interactions, lipophilic interactions, and entropic effects. These uncorrelated terms are usually combined into an overall scoring function by applying linear regression to the descriptor coefficients to fit binding affinity data of a protein–ligand set [116]. Examples of empirical scoring functions include LigScore1, LigScore2, Piecewise Linear Potential (PLP1 and PLP2), Jain and Ludi [117].

Knowledge-based scoring functions calculate the total score as the sum of statistical potentials on interatomic distances from a database of protein–ligand structure complexes [117,118]. The potentials are based on the frequency of occurrence of different atom–atom pair contacts and other typical interactions in large data sets of protein–ligand complexes of known structure [116]. Examples of knowledge-based scoring function are PMF (potential of mean force) and PMF04 [118].

Consensus scoring combines the information obtained from different scoring functions to minimize errors from a single scoring function, thereby improving the probability of finding the

correct solution or 'true' ligand poses [114,116]. Different scoring functions are combined to identify ligands that score high in more than one scoring function, with zero being the minimum value [94].

4.2 Materials and Methods

4.2.1 Virtual High-Throughput Screening of the Bio-Focus library against the *Pf*CDPK4 3D model structure using LibDock

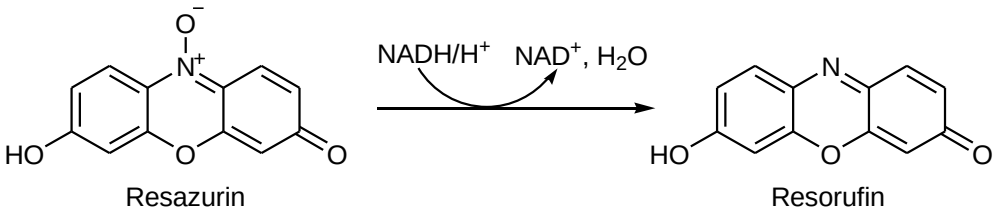
The 83 707 prepared ligands from the Bio-Focus library were screened against the prepared structure of *Pf*CDPK4 using LibDock tools in Discovery Studio 3.1 applying a CHARMm-based force field. LibDock is a docking algorithm designed for *in silico* screening of a large compound database against an identified binding site in a relatively short time. LibDock uses protein site features referred to as hotspots, which are located within the hotspot search volume and may be of two types, polar and apolar. From the generated hotspots, ligand poses are placed into polar and apolar target binding sites. Apolar hotspots are points in the binding site that are favourable for an apolar atom to bind while polar hot spots are points in the binding site that are favourable for a hydrogen bond donor or acceptor to bind [58].

The parameters were set with the conformation method BEST selected and the maximum number of conformations for each ligand set at five (because of the size of Bio-Focus library, selecting a bigger number previously resulted in the output file not opening). A minimization algorithm was not selected, to avoid increased docking computational time as both the ligand library and the *Pf*CDPK4 model were minimized during preparation. The resulting poses docked were scored using the scoring functions PLP1, PLP2, PMF, PMF04, Jain, LigScore1, and LigScore2 (Section 4.1), followed by consensus scoring, and top scoring compounds were selected.

4.2.2 *In vitro* screening of prioritized compounds against *Plasmodium falciparum* strain NF54 late-stage gametocytes using the PrestoBlue™ assay

The antiplasmodial assays against late-stage (IV-V) gametocytes were performed by the Molecular and Biomedical Technologies group of the Council for Scientific and Industrial Research (CSIR), using the following methods:

In vitro gametocidal activity of test samples against the *Plasmodium falciparum* NF54 strain was measured by assessing parasite gametocyte survival after drug exposure using the PrestoBlue™ assay. When cells are viable, they maintain a reducing environment within their cytosol. The PrestoBlue™ cell viability reagent uses that reducing ability to measure quantitatively cell proliferation, and therefore can be used to establish the relative viability of various reagents across many different cell types. The PrestoBlue™ reagent is a resazurin-based solution, which is a blue cell-permeating compound that is virtually non-fluorescent. When added to cells, the reagent reacts with the reducing environment of a viable cell to turn pink in colour and become the highly red fluorescent resorufin:



Resazurin Resorufin

This change was detected by fluorescence measurement at 610 nm using a multi-well spectrophotometer (Tecan Infinite F500). The fluorescence excitation/emission maxima were 550-560/590-615 nm.

Compound inhibitory activity was determined by preparing test samples in parasite culture medium in transparent 96-well flat bottom plates (Greiner Bio-one) at 5 µM concentration (n = 3 for each data point). Parasitized red blood cells were added to a final concentration of 5% haematocrit, 2% gametocytaemia and the plates incubated for 48 hours before proceeding with the PrestoBlue™ assay. Percentage parasite/gametocyte survival in each well was calculated relative to control wells that received no drug (negative control). Dihydroartemisinin (DHA), which exhibits high antimalarial potency against all blood-stages of *P. falciparum*, was used as the positive control.

4.2.3 Docking of known kinase inhibitors against *Pf*CDPK4 3D model structure using CDOCKER

A small set of optimized compounds (36) generated from 13 clinically approved kinase inhibitors was screened for binding to *Pf*CDPK4 using CDOCKER in Discovery Studio 3.5. CDOCKER is a grid-based docking method that employs a CHARMm force field to generate random ligand conformations through high temperature molecular dynamics. The number of random conformations to be generated for each ligand was set to 10 and the molecular

dynamics was calculated with multiple annealing stages of heating and cooling. Energy minimization was performed to optimize each binding conformation generated using a CHARMm force field. Refined poses were scored using scoring functions (PLP1, PLP2, PMF, PMF04, Jain, LigScore1, and LigScore2), and the compounds were then ranked according to their total scores using consensus ranking.

4.2.4 Docking of a small set of Natural and Synthetic Compounds against the *Pf*CDPK4 3D model structure.

The prepared NSC compounds (37) were also screened against the target protein, *Pf*CDPK4, using the CDOCKER protocol with parameter settings the same as described in 4.2.3. The best hits were identified through the application of different scoring functions (PLP1, PLP2, PMF, PMF04, Jain, LigScore1, and LigScore2) followed by consensus scoring.

4.3 Results and Discussion

4.3.1 Virtual High-Throughput Screening of the Bio-Focus Library against the *Pf*CDPK4 3D model structure using LibDock

In silico screening of 83 707 compounds from the Bio-Focus library of compounds resulted in 195 434 poses being produced. The poses were scored using scoring functions including PLP1, PLP2, PMF, PMF04, Jain, LigScore1, and LigScore2. Consensus scoring was employed to find ligands that consistently scored high in more than one scoring function. A total of 26 compounds with the highest consensus scores were prioritized. The selected compounds and their corresponding scores are listed in Table 4 and their structures are shown in Figure 10. The top scoring compounds were identified based on consensus ranking and hydrogen interaction. Compounds with consensus scores of six and five were prioritized for further *in vitro* analysis.

The highest scoring compounds with a consensus score of six were TM1595, TM8696 and TM15669; hence, they can be expected to have potentially the highest binding affinity amongst other selected compounds. Since consensus scoring is a combination of various scoring functions, a consensus score of six results from the combination of six best scoring functions used to score docked complexes. The other twenty-three compounds have a consensus score of five, which means that five scoring functions scored best for those compounds compared to other compounds that were also scored. Appendix B shows how each of these 26 compounds

docks against the *Pf*CDPK4 model. From these figures, it is clear that each compound docks uniquely.

The highest number of amino acid residues involved in hydrogen bonding interaction for the 26 top scoring Bio-Focus-*Pf*CDPK4 complexes are Asp145, Lys8, Glu84, Tyr80 and Lys125 respectively. Figure 11 shows the interactions of some of the amino acid residues of the *Pf*CDPK4 model with atoms of some of the selected compounds.

Table 4: Top scoring compounds and their corresponding scores for the Bio-Focus library against *Pf*CDPK4 docking results.

Compound	LigScore1	LigScore2	PLP1	PLP2	Jain	PMF	PMF04	Consensus
TM1595	4.77	3.59	102.27	103.59	7.29	117.29	66.76	6
TM1993	3.82	3.00	102.85	103.14	6.28	132.98	62.55	5
TM2078	4.28	3.88	90.44	94.28	6.49	132.18	66.39	5
TM2753	-0.04	-0.93	120.34	108.71	7.39	114.52	60.11	5
TM5949	4.30	5.74	111.52	104.54	7.74	90.67	42.12	5
TM6136	3.90	5.46	100.43	93.18	2.87	106.47	50.88	5
TM6139	4.27	4.93	102.28	90.35	5.17	119.36	64.92	5
TM7211	3.15	3.47	99.38	92.95	8.89	112.91	67.88	5
TM8696	3.73	3.67	111.20	97.13	6.69	115.14	74.85	6
TM9336	-0.50	-3.36	102.97	104.00	6.56	127.65	70.04	5
TM13502	4.61	4.75	98.21	95.67	4.31	108.29	42.51	5
TM14954	4.02	4.52	97.09	94.03	6.51	104.40	35.76	5
TM15195	2.68	0.94	100.73	102.02	7.21	129.57	59.23	5
TM15233	4.83	5.29	97.48	90.04	5.68	116.60	62.29	5
TM15235	4.55	4.80	103.18	95.55	6.32	115.46	46.57	5
TM15669	4.17	4.57	100.52	97.84	3.47	126.25	59.32	6
TM15928	3.97	3.38	98.84	98.10	7.33	82.90	65.13	5
TM15930	3.46	4.31	99.54	98.11	4.12	110.19	75.57	5
TM15940	2.18	2.89	99.48	96.33	6.65	114.89	66.61	5
TM16028	3.58	4.11	102.76	99.06	4.99	104.92	77.23	5
TM16037	4.28	3.28	106.62	105.12	7.49	68.36	64.65	5
TM16047	4.28	2.87	103.48	105.94	6.74	94.53	76.42	5
TM16054	4.49	4.41	101.51	101.00	7.73	86.10	64.18	5
TM16065	3.44	2.57	102.01	102.32	8.01	94.92	70.02	5
TM16098	4.00	4.23	97.01	101.90	7.11	102.03	70.72	5
TM19688	4.14	4.04	97.22	96.32	6.43	87.94	66.52	5

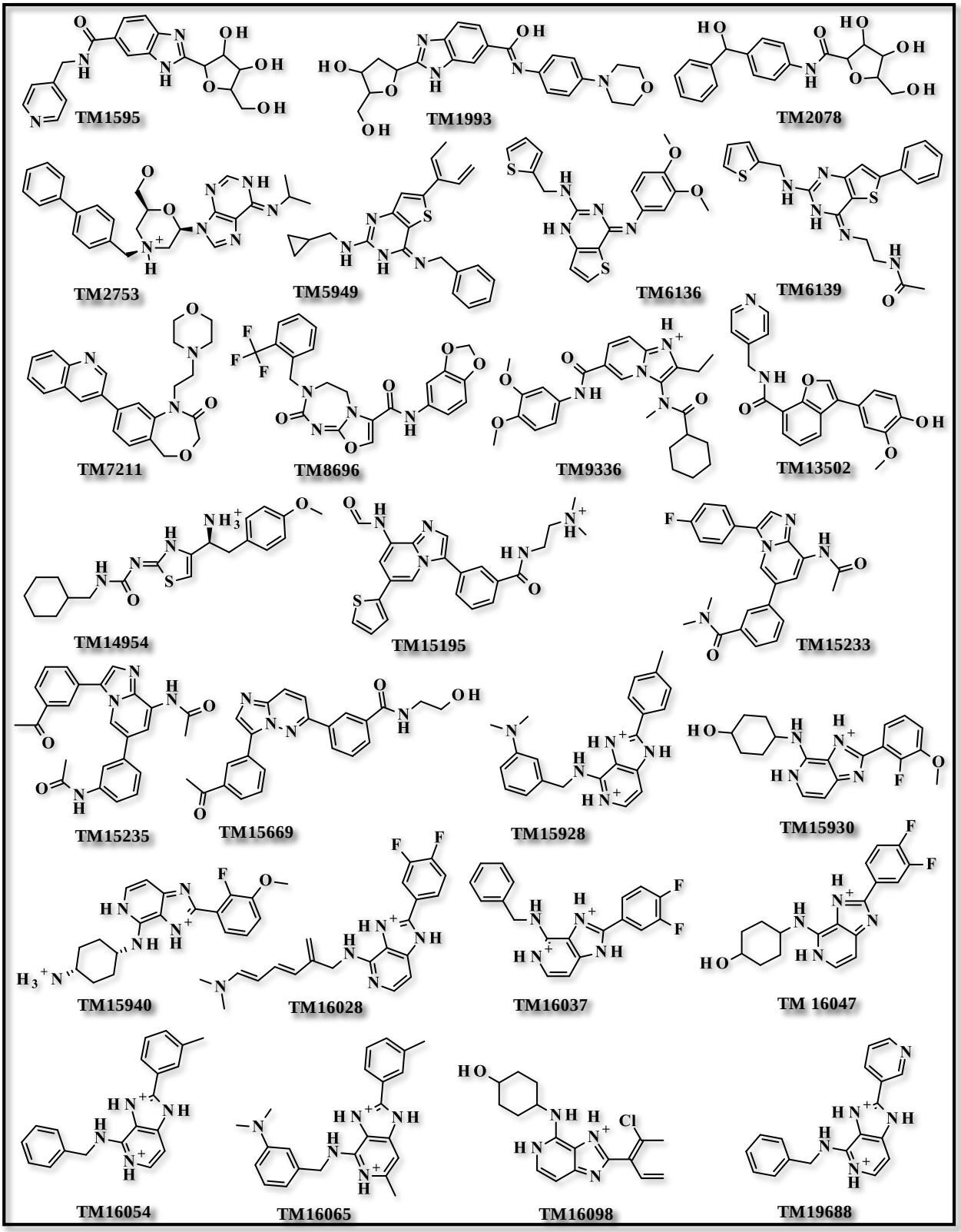


Figure 10: Structures of the selected compounds obtained by virtual screening of the optimized Bio-Focus library against the *Pf*CDPK4 model structure.

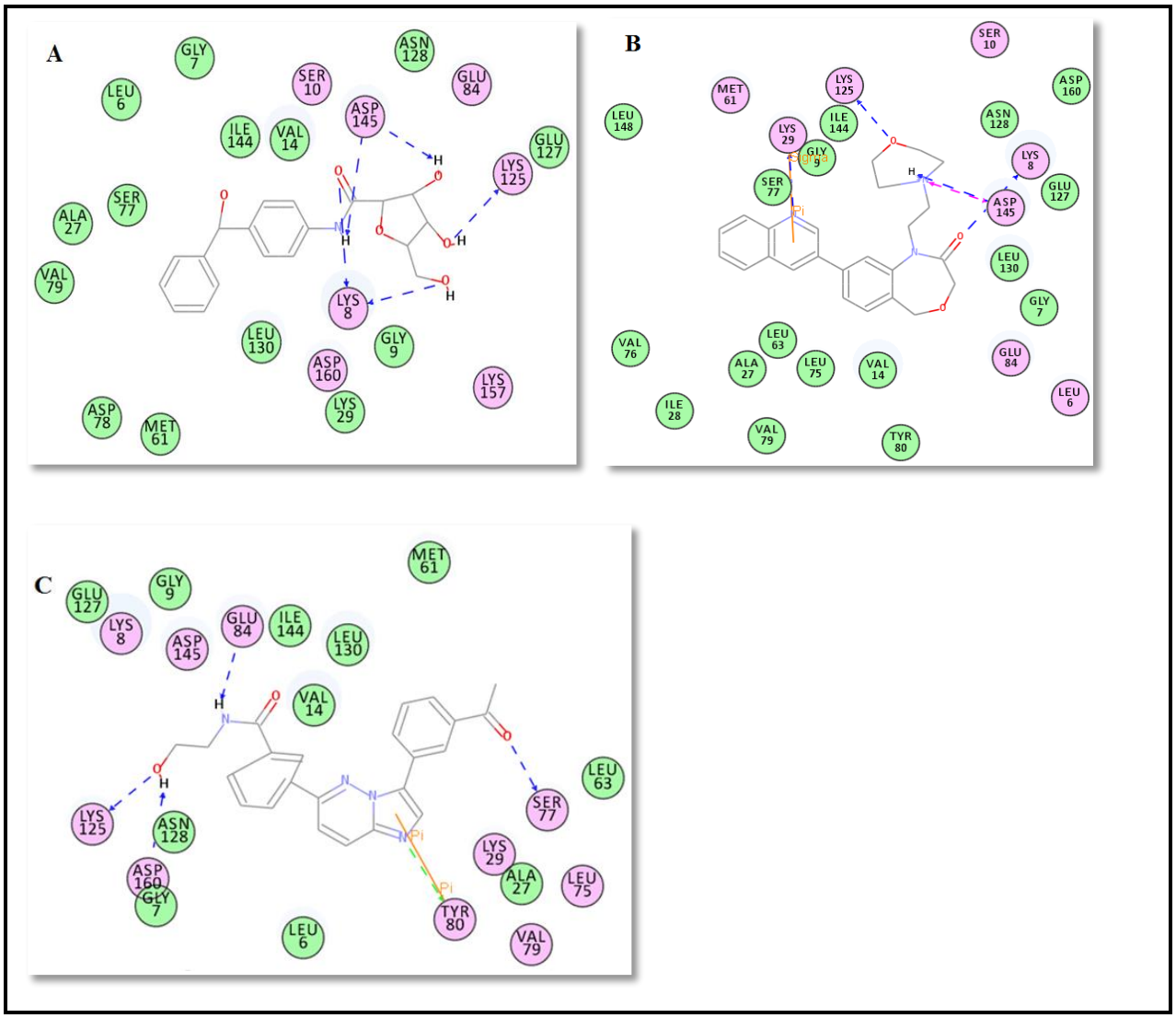


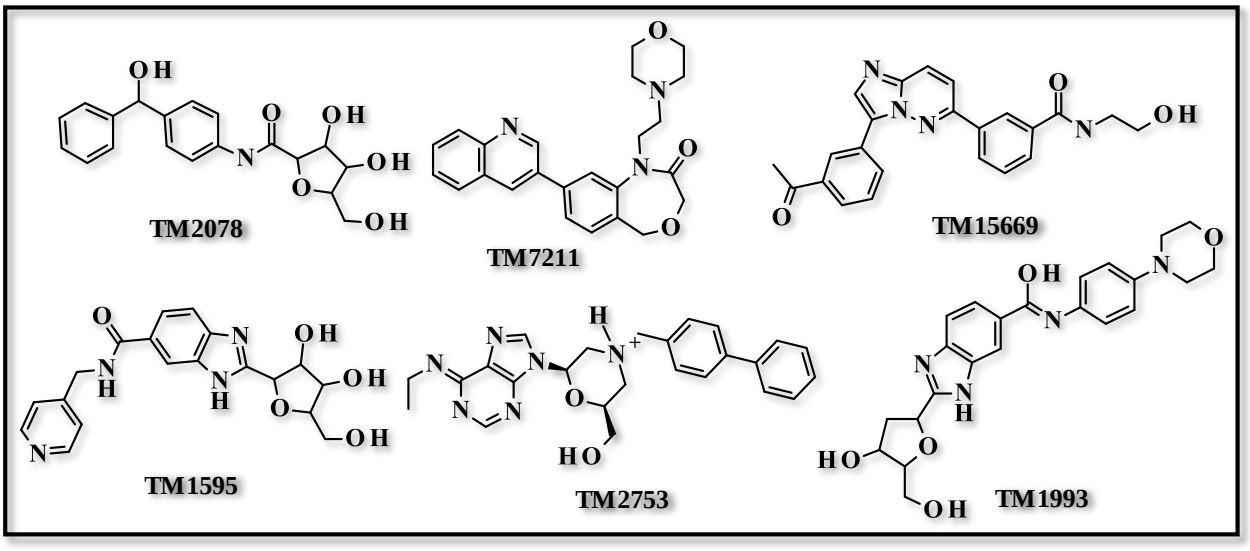
Figure 11: 2-dimensional representation of the interactions of the three compounds with the highest potential binding affinity for the *Pf*CDPK4 active site. Complex A: TM2078-*Pf*CDPK4. Complex B: TM7211-*Pf*CDPK4. Complex C: TM15669-*Pf*CDPK4. The blue and green arrows indicate hydrogen-bonding interactions and the orange arrows indicate Pi-Sigma and Pi-Pi interactions.

TM2078 is predicted to make hydrogen bond interactions with Lys8, Lys125 and Asp145 amino acid residues of *Pf*CDPK4 (Figure 11A). TM7211 is predicted to form hydrogen bond interactions with Asp145, Lys8 and Lys125, as well as attractive charge (ionic bond) with Glu84 and Asp145 (Figure 11B). TM15669 is predicted to form hydrogen bonds with Ser77, Tyr80, Lys125, Glu84 and Asp160, as well as Pi-anion bond with Glu84 (Figure 11C). The DXR inhibitor that was bound to the crystal structure of *Tg*CDPK1 shows hydrogen bond interaction with Glu112 and Tyr114, as well as hydrophobic (Pi-Sigma) interactions with Val48, Met95 and Leu164. Predicted hydrogen bond interaction for compound TM15669 with

Glu84 and Tyr80 amino acid residues was also observed in the crystal template structure of *Tg*CDPK1 in complex with inhibitor DXR where Tyr114 and Glu112 are involved in the hydrogen bond interaction. Similar prediction was also observed where a substituent of bumped kinase inhibitor-1 was predicted to form a hydrogen bond with the Glu184 amino acid residue of *Pf*CDPK4 which was observed before in crystal structures of *Tg*CDPK1 in complex with bumped kinase inhibitors (BKIs) [32].

4.3.2 *In vitro* screening of prioritized compounds against Plasmodium falciparum NF54 strain at late gametocytes stage using PrestoBlue™ assay

In vitro screening was carried out for all the prioritized compounds for activity against the *Plasmodium falciparum* NF54 strain, assayed at a single concentration of 5 µM. The compounds and their corresponding percentage inhibitory values are shown in Figure 12. As a general rule for reference activities, when interpreting data obtained from the single concentration assay determined at 5 µM, compounds can be classified as inactive (less than 20% gametocyte inhibition), marginally active (gametocyte inhibition greater than 50%) and good activity (% gametocyte inhibition greater than 70%). Of the 26 compounds that were screened, only three compounds (TM2078, TM7211, and TM15669, shown below) exhibited inhibitory values close to marginally active/active and the rest of the compounds are classified as inactive/poorly active. The percentage inhibition for these three compounds is: TM2078 - 46%, TM7211 – 47% and TM15669 – 42%. The *in vitro* screening results also show that three other compounds were found above 20% inhibition and their percentage inhibition is as follows: TM1595 – 37%, TM2753 – 33% and TM1993 – 27%.



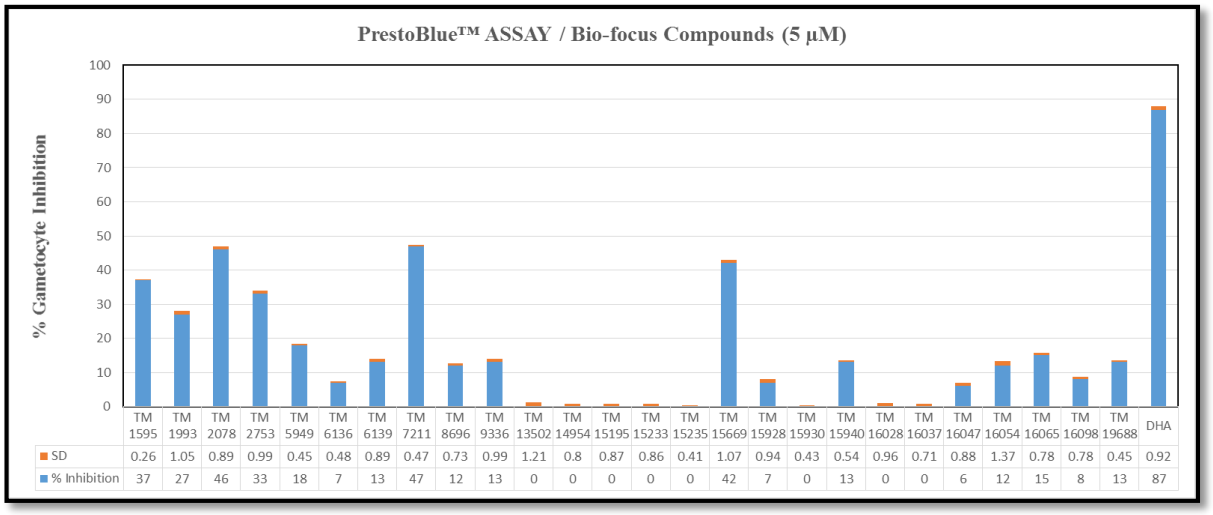


Figure 12: *In vitro* screening results for 26 selected compounds against *Plasmodium falciparum* malaria strain NF54.

4.3.3 Docking of known kinase inhibitors against *Pf*CDPK4 3D model structure using CDOCKER

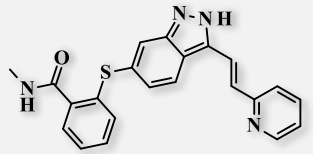
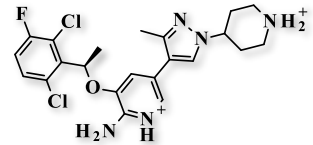
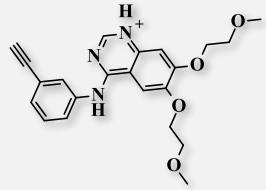
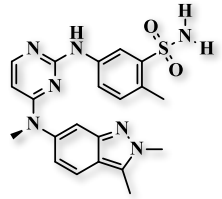
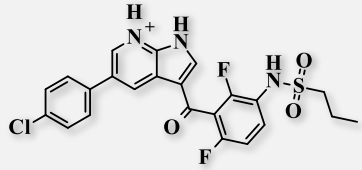
From the 36 kinase inhibitors that were docked against PCDPK4, 175 poses from 23 compounds were docked while 13 compounds failed. Scoring was applied to all 175 poses to predict the binding affinity of the poses and to identify highly active compounds. After consensus ranking, five compounds (6450551, 11626560, 176870, 10113978 and 42611257) were selected according to their predicted binding affinity; they are listed in Table 6. The 2-dimensional representations of the five selected compounds are shown in Figure 13. Kinase inhibitors screened against *Pf*CDPK4 in this study are FDA (Food and Drug Administration) approved and are listed in Table 5 along with their specific target(s) and their clinical applications [119]. The kinase inhibitors were docked to find out if they could interact with *Pf*CDPK4 since they are active against other kinase targets. Unfortunately, due to time constraints and availability of compounds no *in vitro* screening was done for the kinase inhibitors that were found to have high binding affinities from the docking results. However, *in silico* screening did predict that some of the kinase inhibitors docked would have potential for further investigation as possible drugs for malaria, targeting *Pf*CDPK4. The strategy of using approved drug compounds can be very helpful since they provide information about their targets and scaffolding. These compounds can be used to validate other homologous targets that

may be helpful in drug discovery for other diseases. The greatest advantage with approved drug compounds is when it comes to toxicity, as compounds will have been tested before.

Table 5: List of FDA approved protein kinase inhibitors (www.brimr.org/PKIs/htm).

Name	Target	Indication	Company	FDA approval
Axitinib	VEGFR1/VEGFR2/VEGFR3/PDGFRB/c-KIT	RCC	Pfizer	2012
Crizotinib	ALK/Met	NSCLC with Alk mutation	Pfizer	2011
Dasatinib	BCR-Abl, Src, Lck, Yes, Fyn, Kit, EphA2, and PDGFRβ	CML	Bristol-Myers Squibb	2006
Erlotinib	Erb1, EGFR	NSCLC and pancreatic cancer	Genentech/Roche	2004
Gefitinib	EGFR	NSCLC	AstraZeneca	2003
Imatinib	BCR-Abl, Kit, and PDGFR	CML, ALL, aggressive systemic mastocytosis, GIST	Novartis	2001
Lapatinib	EGFR and ErbB2	Breast cancer	GSK	2007
Nilotinib	BCR-Abl, PDGFR	CML	Novartis	2007
Pazopanib	VEGFR2/PDGFR/c-kit	RCC, soft tissue sarcomas	GlaxoSmithKline	2009
Ruxolitinib	Eph2A, JAK1/2	Myelofibrosis and PV	Incyte	2011
Sorafenib	C-Raf, B-Raf, B-Raf (V600E), Kit, Flt3, RET, VEGFR1/2/3, and PDGFRα/β	Hepatocellular carcinoma, RCC, DTC (updated 23 Nov 2013)	Onyx/Bayer	2005
Sunitinib	PDGFRα/β, VEGFR1/2/3, Kit, Flt3, CSF-1R, and RET	RCC, GIST, pancreatic neuroendocrine tumors	SUGEN/*Pfizer*	2006
Vemurafenib	A/B/C-Raf and B-Raf (V600E)	Melanoma with $BRAF^{V600E}$ mutation	Hoffmann La Roche	2011

Table 6: Selected kinase inhibitors with their corresponding scores and structures.

Compound	LigScore1	LigScore2	PLP1	PLP2	Jain	PMF	PMF04	Consensus Score	Structure
6450551	3.22	5.51	90.79	79.70	3.67	99.60	40.30	7	
11626560	3.43	6.16	96.62	82.45	3.88	82.74	51.45	5	
176870	3.34	5.41	85.66	79.34	2.44	65.39	32.29	5	
10113978	3.44	5.49	94.38	86.11	6.01	101.59	50.07	6	
42611257	4.37	5.74	81.07	79.53	5.22	87.50	43.14	5	

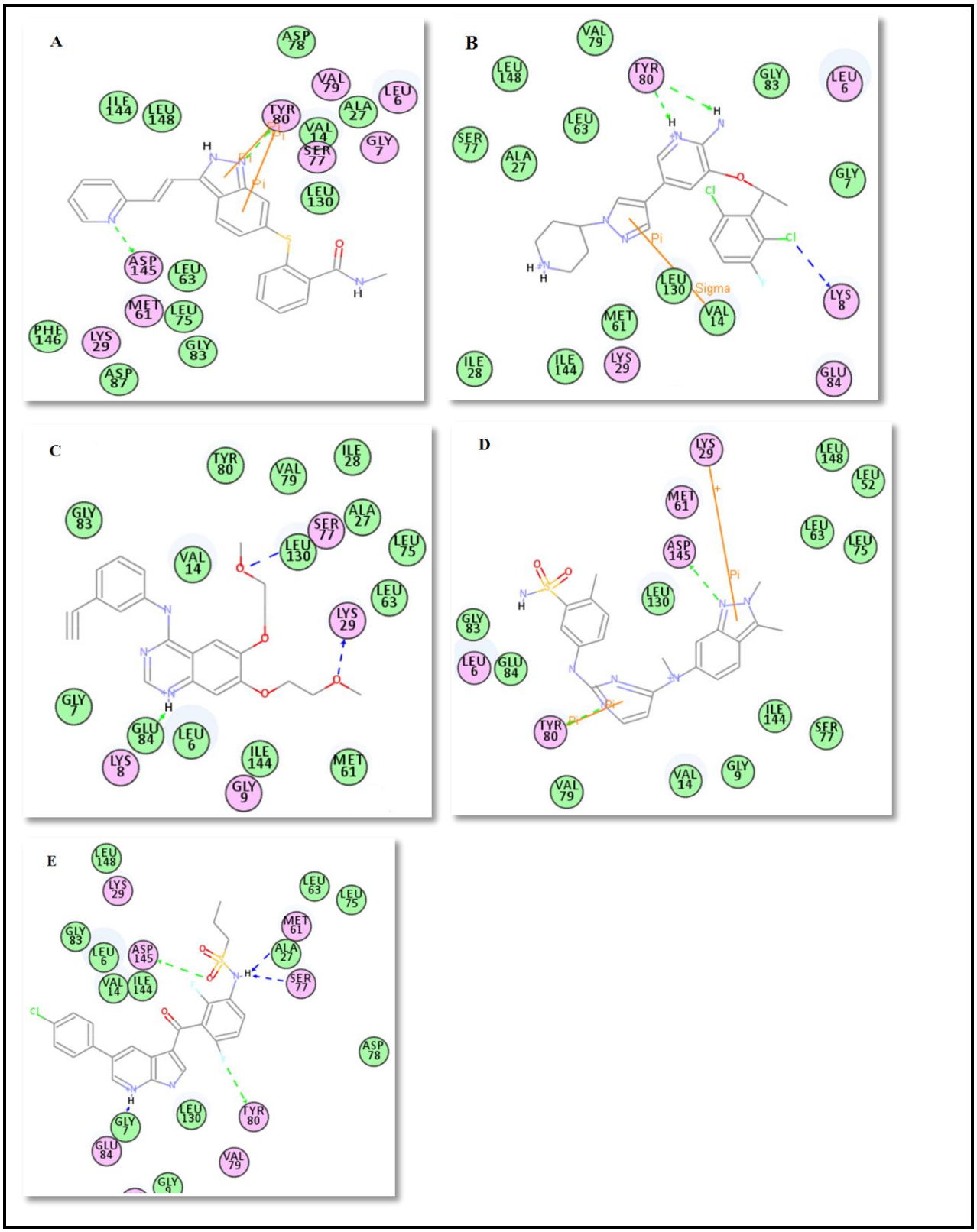
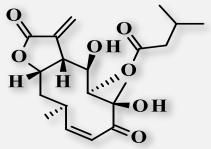
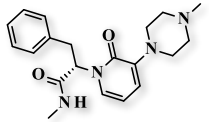
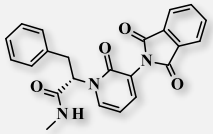
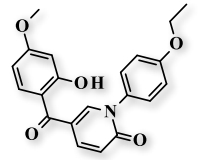
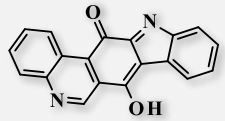
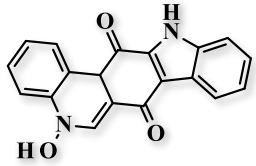


Figure 13: Kinase inhibitors versus *Pf*CDPK4 in 2-dimensional representation for hydrogen bonding interactions of selected compounds. A: 6450551. B: 11626560. C: 176870. D: 10113978. E: 42611257. The blue and green arrows indicate hydrogen bond interactions and the orange arrow indicates Pi-Sigma and Pi-Pi interactions.

4.3.4 Docking of a small set of Natural and Synthetic Compounds against the *Pf*CDPK4 3D model structure.

The docking of 37 optimized ligands from the NSC library against *Pf*CDPK4 resulted in 308 poses from 28 compounds. From the docked hits, six compounds were selected based on the scoring and consensus ranking (Table 7 and Figure 14). Neurolenin C [120], Calothrixin A and B [121] are amongst the hits with the highest binding affinity. These are isolated natural compounds and they exhibited antimalarial activity in previous *in vitro* screening. The targets for these compounds remain unknown. This study has highlighted that natural compounds can also be used to search for potential targets. Since docking of these compounds has shown them to bind with *Pf*CDPK4, further investigation is required to establish if indeed *Pf*CDPK4 can be their potential target.

Table 7: NSC selected compounds predicted to have the highest binding affinity with their corresponding scores and structures.

Compound	LigScore1	LigScore2	PLP1	PLP2	Jain	PMF	PMF04	Consensus Score	Structure
NeuroC	3.12	4.72	76.89	72.40	5.49	77.17	14.59	5	
Compd5	2.75	5.56	85.65	77.57	4.60	54.23	14.98	6	
Compd4	3.56	5.76	94.67	83.01	3.85	91.67	26.07	7	
Compd3	2.56	5.22	76.51	69.47	2.13	84.06	29.38	5	
CaloB	2.95	5.44	77.24	71.24	3.02	90.05	38.30	5	
CaloA	4.17	5.72	83.52	77.28	3.68	71.03	35.18	6	

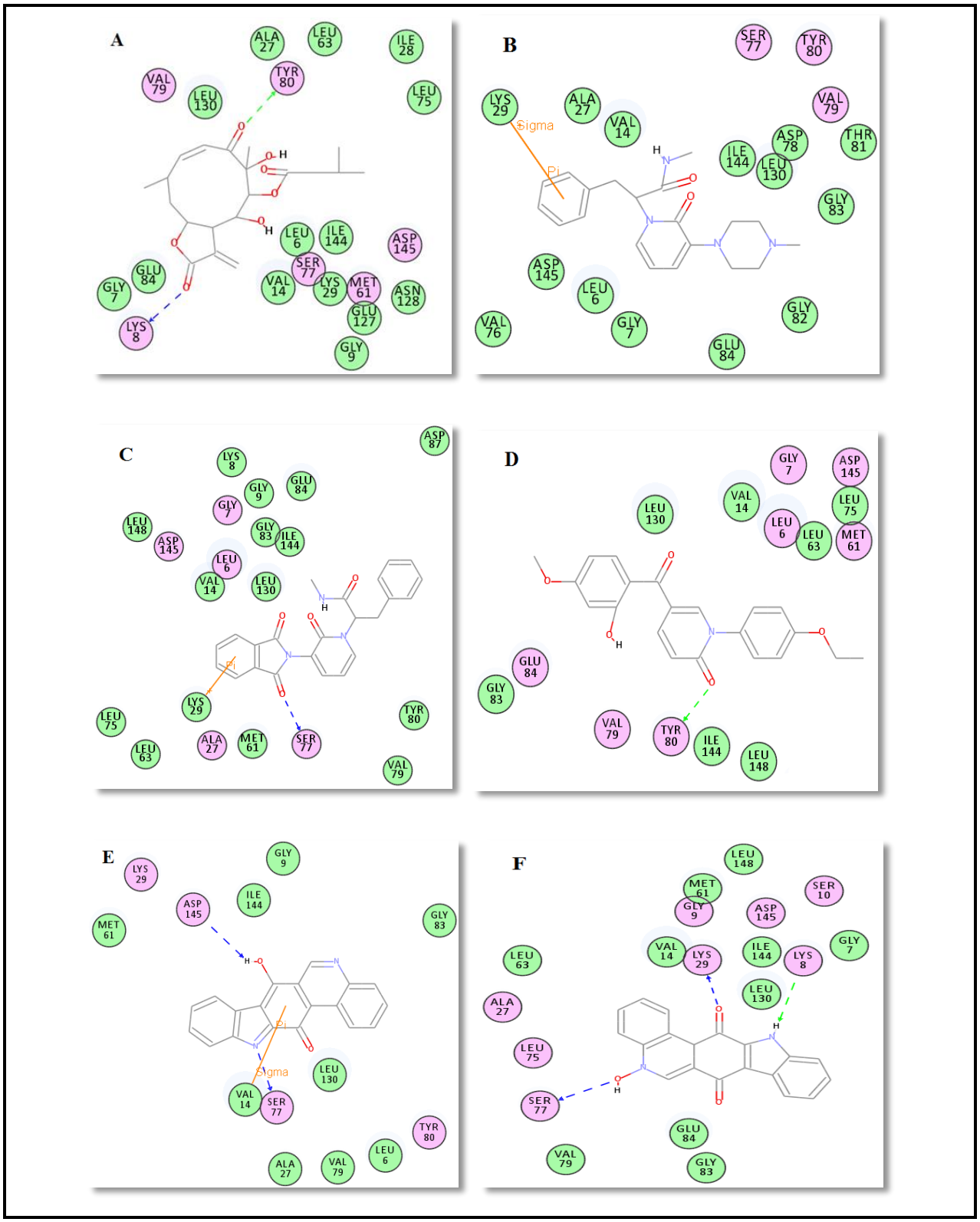


Figure 14: Two-dimensional representation for selected hits from NSC versus *Pf*CDPK4 docking results. A: NeuroC. B: Compd5. C: Compd4. D: Compd3. E: CaloB. F: CaloA. The blue and green arrows indicate hydrogen bond interactions and the orange arrow indicate Pi-Sigma and Pi-Pi interactions.

CHAPTER 5

CONCLUDING REMARKS

It is possible to implement *in silico* structure-based screening, regardless of the non-availability of certain target structures in protein structure databases. Homology modelling is considered the most reliable and accurate method for predicting the 3D structures of most proteins in the absence of experimentally determined structures, and aims to predict accurately the structure of a protein whose crystal or NMR structure has not yet been solved. The technique makes use of the target protein's amino acid sequence to create a 3-dimensional model with molecular detail and accuracy that approaches the results of experimentally solved structures.

Since the structure of *Pf*CDPK4 was not available in protein structure databases and has not yet been solved experimentally, homology modelling was carried out using the X-ray crystallographic structure of *Tg*CDPK1 (PDB code: 3ma6-b*)* as a template to generate a structure of *Pf*CDPK4 that is suitable for drug discovery applications. The quality and the reliability of the model target (*Pf*CDPK4) depends highly on the degree of similarity of the sequence between the template – *Tg*CDPK1 (3ma6 b) – and the target (*Pf*CDPK4). With 79% sequence identity, it was clear that 3ma6 is the best template amongst other homologous templates from the BLAST Search hit list to build the 3-dimensional structure of *Pf*CDPK4. The alignment between *Pf*CDPK4 and *Tg*CDPK1 showed 78 % sequence identity and 90 % sequence similarity suggesting that the model built with this alignment is acceptable.

It is important to have knowledge of the 3D structure of the target protein because it is a prerequisite for structure-based drug design. Homology modelling is a stepwise protocol and hence it is important to carry out each step appropriately to avoid errors that may compromise the quality for the resulting model. It is important to select the template carefully as the quality of the model structure relies on the template. A homology-modelled structure requires validation to make sure that the model is of good quality and hence can be used for further investigations. The homology model structure of *Pf*CDPK4 modelled using a currently available homologous structure (*Tg*CDPK1) as template was validated with good stereochemical quality. The non-availability of the 3-dimensional PDB structure of *Pf*CDPK4 therefore has not been an obstacle to this study, since homology modelling made it available.

For molecular docking to be successful, several steps were followed. Those steps are homology model building, binding site identification from cavities, preparation and docking of ligands, which was followed by scoring and analysis. After building the target or receptor structure, it is important to identify the binding site, which provides guidance as to where the ligand is likely to bind. For the *Pf*CDPK4 model structure, the binding site was identified from structure cavities. Seven binding sites were identified, but the correct binding site of interest was selected aided by the existing knowledge about kinases binding ATP.

The compounds docked were prepared by first filtering them to make sure that compounds that are not suitable to be drugs are eliminated. The Lipinski Rule of Five [99] criteria were used so that no ligand with more than 5 hydrogen bond donors, molecular weight greater than 500, Log*P* over 5 or more than 10 hydrogen acceptors was further used for docking. From the filtered libraries (Bio-Focus, NSC) of compounds, a notable number of compounds did not pass the Lipinski Rule of Five criteria, suggesting that they do not have good drug-like properties.

In this study, the 26 compounds with the highest binding affinity were selected by *in silico* screening. From these 26 priority compounds, *in vitro* screening confirmed that three compounds exhibit some activity against *Plasmodium falciparum* strain NF54. Some clinically approved kinase inhibitors that were screened against *Pf*CDPK4 have also been shown to interact with the target, which suggests that five selected compounds can be considered for further investigation. Some compounds from the NSC library also bind against *Pf*CDPK4 and six compounds with the highest binding affinity were prioritized. The overall results suggest that *Pf*CDPK4 can be used as target for searching for novel lead compounds for the therapeutic treatment of malaria.

Having identified several promising hits (Figure 18), this study should be followed up with additional work pursuing the following objectives:

- To identify a set of hits most suitable for potential further development into leads for the treatment of malaria.
- To modify the identified hits through medicinal chemistry and organic synthesis.
- To systematically examine more anti-malarial natural compounds for binding to *Pf*CDPK4.
- To screen Bio-Focus hits against other life-circle stages of malaria parasite.

- To screen known kinase inhibitors of interest against other life-circle stages of malaria parasite.

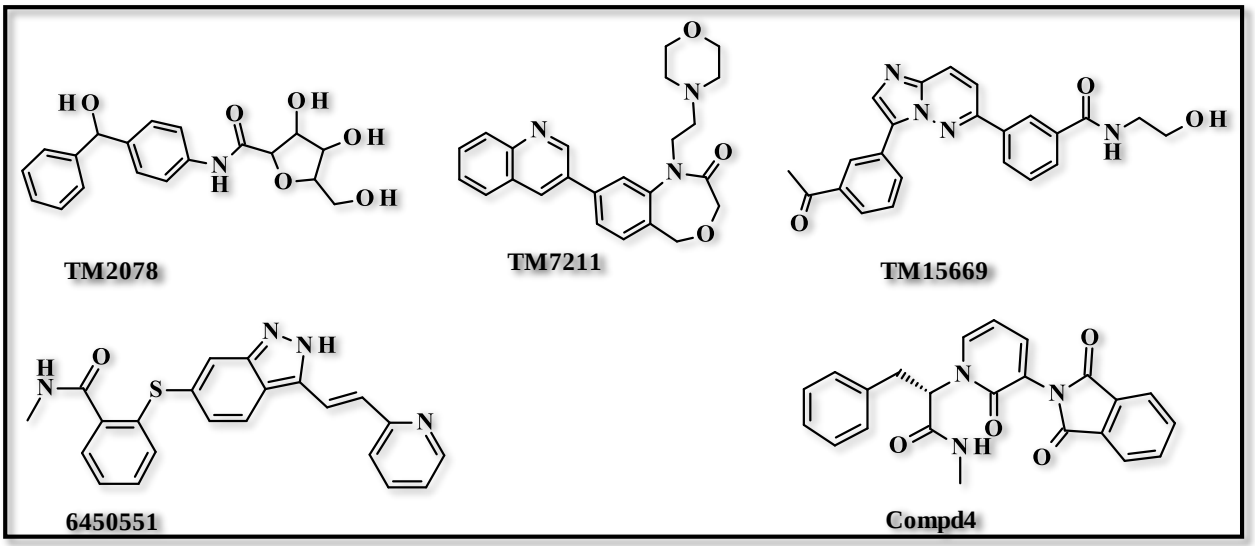


Figure 15: A sample of promising hits produced by this study.

References

[1] M. A. Biamonte, J. Wanner, and K. G. Le Roch, "Recent advances in malaria drug discovery," *Bioorg Med Chem Lett*, 23, pp. 2829-2843, 2013.

[2] V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R. Bywater, and N. Foloppe, "Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery," *Comput Struct Biotechnol J*, 5, pp. 1-14, 2013.

[3] D. Dorin-Semblat, T. G. Carvalho, M.-P. Nivez, J. Halbert, P. Poullet, J.-P. Semblat, *et al.*, "An atypical cyclin-dependent kinase controls *Plasmodium falciparum* proliferation rate," *Kinome*, 1, pp. 4-16, 2013.

[4] WHO (World Health Organization), "World Malaria Report," 2013.

[5] R. S. Vidadala, K. K. Ojo, S. M. Johnson, Z. Zhang, S. E. Leonard, A. Mitra, *et al.*, "Development of potent and selective *Plasmodium falciparum* calcium-dependent protein kinase 4 (*Pf*CDPK4) inhibitors that block the transmission of malaria to mosquitoes," *Eur J Med Chem*, 74, pp. 562-573, 2014.

[6] E. Y. Klein, "Antimalarial drug resistance: a review of the biology and strategies to delay emergence and spread," *Int J Antimicrob Agents*, 41, pp. 311-317, 2013.

[7] I. Petersen, R. Eastman, and M. Lanzer, "Drug-resistant malaria: Molecular mechanisms and implications for public health," *FEBS Lett*, 585, pp. 1551-1562, 2011.

[8] C. J. L. Murray, L. C. Rosenfeld, S. S. Lim, K. G. Andrews, K. J. Foreman, D. Haring, *et al.*, "Global malaria mortality between 1980 and 2010: A systematic analysis," *Lancet*, 379, pp. 413-431, 2012.

[9] WHO (World Health Organization), "World Malaria Report 2012 Fact Sheet," 2012.

[10] J. Held, A. Kreidenweiss, and B. Mordmüller, "Novel approaches in antimalarial drug discovery," *Expert Opin Drug Discov*, 8, pp. 1325-1337, 2013.

[11] R. Tewari, U. Straschil, A. Bateman, U. Bohme, I. Cherevach, P. Gong, *et al.*, "The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission," *Cell Host Microbe*, 8, pp. 377-87, 2010.

[12] H.-H. Chang, E. L. Moss, D. J. Park, D. Ndiaye, S. Mboup, S. K. Volkman, *et al.*, "Malaria life cycle intensifies both natural selection and random genetic drift," *Proc Natl Acad Sci*, 110, pp. 20129-20134, 2013.

[13] E. R. Derbyshire, M. M. Mota, and J. Clardy, "The Next Opportunity in Anti-Malaria Drug Discovery: The Liver Stage," *PLoS Pathog*, 7, p. e1002178, 2011.

[14] L. H. Miller, D. I. Baruch, K. Marsh, and O. K. Doumbo, "The pathogenic basis of malaria," *Nature*, 415, pp. 673-679, 2002.

[15] P. Ward, L. Equinet, J. Packer, and C. Doerig, "Protein kinases of the human malaria parasite *Plasmodium* falciparum: the kinome of a divergent eukaryote," *BMC Genomics*, 5, p. 79, 2004.

[16] S. P. Shekinah and M. Rajadurai, "A Study of *in silico* Drug Docking for Triosephosphate Isomerase in *Plasmodium falciparum*," *Advanced BioTech*, 7, pp. 10-13, 2008.

[17] L. H. Miller, M. F. Good, and G. Milon, "Malaria pathogenesis," *Science*, 264, pp. 1878-1883, 1994.

[18] A. Heddini, "Malaria pathogenesis: a jigsaw with an increasing number of pieces," *Int J Parasitol*, 32, pp. 1587-1598, 2002.

[19] WHO (World Health Organization), "World Malaria Report 2011," 2011.

[20] WHO (World Health Organization), "Global plan for artemisinin resistance containment (GPARC)," 2011.

[21] M. Schlitzer, "Antimalarial drugs - what is in use and what is in the pipeline," *Arch Pharm (Weinheim, Ger)*, 341, pp. 149-163, 2008.

[22] D. C. Lim, B. M. Cooke, C. Doerig, and J. P. J. Saeij, "Toxoplasma and *Plasmodium* protein kinases: Roles in invasion and host cell remodelling," *Intl J Parasitol*, 42, pp. 21-32, 2012.

[23] T. Anastassiadis, S. W. Deacon, K. Devarajan, H. Ma, and J. R. Peterson, "Comprehensive assay of kinase catalytic activity reveals features of kinase," *Nat Biotechnol*, 29, pp. 1039-1045, 2011.

[24] R. Ranjan, A. Ahmed, S. Gourinath, and P. Sharma, "Dissection of mechanisms involved in the regulation of *Plasmodium falciparum*," *J Biol Chem*, 284, pp. 15267-15276, 2009.

[25] T. M. Chapman, S. A. Osborne, N. Bouloc, J. M. Large, C. Wallace, K. Birchall, *et al.*, "Substituted imidazopyridazines are potent and selective inhibitors of *Plasmodium falciparum* calcium-dependent protein kinase 1 (*Pf*CDPK1)," *Bioorg Med Chem Lett*, 23, pp. 3064-3069, 2013.

[26] L. Solyakov, J. Halbert, M. M. Alam, J. P. Semblat, D. Dorin-Semblat, L. Reininger, *et al.*, "Global kinomic and phospho-proteomic analyses of the human malaria parasite," *Nat Commun*, 2, p. 565, 2011.

[27] S. K. Middha, A. K. Goyal, S. A. Faizan, N. Sanghamitra, B. C. Basistha, and T. Usha, "*In silico*-based combinatorial pharmacophore modelling and docking studies of GSK-3beta and GK inhibitors of Hippophae," *J Biosci*, 38, pp. 805-814, 2013.

[28] K. K. Ojo, C. *Pf*ander, N. R. Mueller, C. Burstroem, E. T. Larson, C. M. Bryan, *et al.*, "Transmission of malaria to mosquitoes blocked by bumped kinase inhibitors," *J Clin Invest*, 122, pp. 2301-2305, 2012.

[29] A. Bansal, S. Singh, K. R. More, D. Hans, K. Nangalia, M. Yogavel, *et al.*, "Characterization of *Plasmodium falciparum* calcium-dependent protein kinase 1 (*Pf*CDPK1) and its role in microneme secretion during erythrocyte invasion," *J Biol Chem*, 288, pp. 1590-602, 2013.

[30] R. Ranjan, A. Ahmed, S. Gourinath, and P. Sharma, "Dissection of Mechanisms Involved in the Regulation of *Plasmodium falciparum* Calcium-dependent Protein Kinase 4," *J Biol Chem*, 284, pp. 15267-15276, 2009.

[31] O. Billker, S. Dechamps, R. Tewari, G. Wenig, B. Franke-Fayard, and V. Brinkmann, "Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite," *Cell*, 117, pp. 503-514, 2004.

[32] K. K. Ojo, R. T. Eastman, R. Vidadala, Z. Zhang, K. L. Rivas, R. Choi, *et al.*, "A specific inhibitor of *Pf*CDPK4 blocks malaria transmission: chemical-genetic validation," *J Infect Dis*, 209, pp. 275-84, 2014.

[33] F. Mojab, "Antimalarial natural products: a review," *Avicenna J Phytomed*, 2, pp. 52-62, 2012.

[34] H. Ginsburg and E. Deharo, "A call for using natural compounds in the development of new antimalarial treatments - an introduction," *Malaria Journal*, 10 Suppl 1, pp. 1-7, 2011.

[35] E. Guantai and K. Chibale, "How can natural products serve as a viable source of lead compounds for the development of new/novel anti-malarials?," *Malaria Journal*, 10 Suppl 1, p. S2, 2011.

[36] Y. Pevzner, D. N. Santiago, J. L. von Salm, R. S. Metcalf, K. G. Daniel, L. Calcul, *et al.*, "Virtual target screening to rapidly identify potential protein targets ofnatural products in drug discovery," *AIMS Molecular Science* 1, pp. 81-98, 2014.

[37] H. Zaid, J. Raiyn, A. Nasser, B. Saad, and A. Rayan, "Physicochemical Properties of Natural Based Products versus Synthetic Chemicals," *Open Nutraceuticals J*, 3, pp. 194-202, 2010.

[38] F. Ntie-Kang, P. A. Onguéné, L. L. Lifongo, J. C. Ndom, W. Sippl, and L. M. a. Mbaze, "The potential of anti-malarial compounds derived from African medicinal plants, part II: a pharmacological evaluation of non-alkaloids and non-terpenoids," *Malaria Journal*, 13, pp. 81-81, 2014.

[39] J. L. Medina-Franco, "Advances in computational approaches for drug discovery based on natural products," *Revista Latinoamericana de Química*, 41, pp. 95-110, 2013.

[40] V. Rao and K. Srinivas, "Modern drug discovery process: An *in silico* approach," *Journal of Bioinformatics and Sequence Analysis*, 2, pp. 89-94, 2011.

[41] S. Mandal, M. n. Moudgil, and S. K. Mandal, "Rational drug design," *Eur J Pharmacol*, 625, pp. 90-100, 2009.

[42] B. K. Shoichet, "Virtual screening of chemical libraries," *Nature*, 432, pp. 862-865, 2004.

[43] G. Klebe, "Virtual ligand screening: strategies, perspectives and limitations," *Drug Discovery Today*, 11, pp. 580-594, 2006.

[44] V. Vyas, A. Jain, A. Jain, and A. Gupta, "Virtual screening: a fast tool for drug design," *Sci Pharm*, 76, pp. 333-360, 2008.

[45] E. M. Krovat, T. Steindl, and T. Langer, "Recent Advances in Docking and Scoring," *Curr Comput-Aided Drug Des*, 1, pp. 93-102, 2005.

[46] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular Docking: A powerful approach for structure-based drug discovery," *Curr Comput-Aided Drug Des*, 7, pp. 146-157, 2011.

[47] A. Lavecchia and C. Di Giovanni, "Virtual screening strategies in drug discovery: a critical review," *Curr Med Chem*, 20, pp. 2839-2860, 2013.

[48] G. Melagraki and A. Afantitis, "Ligand and structure based virtual screening strategies for hit-finding and optimization of hepatitis C virus (HCV) inhibitors," *Curr Med Chem*, 18, pp. 2612-2619, 2011.

[49] A. Hamza, N. N. Wei, and C. G. Zhan, "Ligand-based virtual screening approach using a new scoring function," *J Chem Inf Model*, 52, pp. 963-74, 2012.

[50] P. Aparoy, K. Kumar Reddy, and P. Reddanna, "Structure and Ligand Based Drug Design Strategies in the Development of Novel 5- LOX Inhibitors," *Curr Med Chem*, 19, pp. 3763-3778, 2012.

[51] S. Jamal, V. Periwal, and V. Scaria, "Predictive modeling of anti-malarial molecules inhibiting apicoplast formation," *BMC Bioinformatics*, 14, p. 55, 2013.

[52] WHO (World Health Organization), "Guidelines for the treatment of malaria," 2010.

[53] V. Zoete, A. Grosdidier, and O. Michielin, "Docking, virtual high throughput screening and *in silico* fragment-based drug design," *Journal for Cellular and Molecular Medicine*, 13, pp. 238-248, 2009.

[54] G. C. Terstappen and A. Reggiani, "*In silico* research in drug discovery," *Trends Pharmacol Sci*, 22, pp. 23-26, 2001.

[55] A. O. T. Bishop, T. A. P. d. Beer, and F. Joubert, "Protein homology modelling and its use in South Africa," *South African Journal of Science*, 104, pp. 2-6, 2008.

[56] C. N. Cavasotto and S. S. Phatak, "Homology modeling in drug discovery: current trends and applications," *Drug Discovery Today*, 14, pp. 676-683, 2009.

[57] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res*, 25, pp. 3389-3402, 1997.

[58] H. Yu, H. Jin, L. Sun, L. Zhang, G. Sun, Z. Wang, *et al.*, "Toll-like receptor 7 agonists: chemical feature based pharmacophore identification and molecular docking studies," *PLoS One*, 8, pp. 1-11, 2013.

[59] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-y. Shen,U. Pieper, and A. Sali, "Comparative Protein Structure Modeling Using Modeller," *Curr Protoc Bioinformatics*, pp. 5.6.1-5.6.30, 2002.

[60] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, 22, pp. 195-201, 2006.

[61] K. Arnold, F. Kiefer, J. Kopp, J. D. Battey, M. Podvinec, J. Westbrook, *et al.*, "The Protein Model Portal," *J Struct Funct Genomics*, 10, pp. 1-8, 2009.

[62] S. Kalyaanamoorthy and Y. P. Chen, "Modelling and enhanced molecular dynamics to steer structure-based drug discovery," *Prog Biophys Mol Biol*, 114, pp. 123-36, 2014.

[63] O. Elbegdorj, R. B. Westkaemper, and Y. Zhang, "A homology modeling study toward the understanding of three-dimensional structure and putative pharmacological profile of the G-protein coupled receptor GPR55," *J Mol Graphics Modell*, 39, pp. 50-60, 2013.

[64] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J Comput Chem*, 4, pp. 187-217, 1983.

[65] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *J Comput Chem*, 25, pp. 1157-74, 2004.

[66] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, *et al.*, "The GROMOS Biomolecular Simulation Program Package," *J Phys Chem A*, 103, pp. 3596-3607, 1999.

[67] S.-W. Lee, N.-R. Lee, J.-H. Lee, and T.-J. Oh, "Homology Modeling and Molecular Docking Analysis of Streptomyces peucetius CYP125A4 as C26 Monooxygenase " *Bull Korean Chem Soc*, 33, pp. 1885-1889, 2012.

[68] J. Durrant and J. A. McCammon, "Molecular dynamics simulations and drug discovery," *BMC Biology*, 9, p. 71, 2011.

[69] U. Consortium, "Reorganizing the protein space at the Universal Protein Resource (UniProt)," *Nucleic Acids Res*, 40, pp. D71-5, 2012.

[70] N. Deshpande, K. J. Addess, W. F. Bluhm, J. C. Merino-Ott, W. Townsend-Merino, Q. Zhang, *et al.*, "The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema," *Nucleic Acids Res*, 33, pp. D233-D237, 2005.

[71] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Res*, 22, pp. 4673-80, 1994.

[72] A. Šali, L. Potterton, F. Yuan, H. van Vlijmen, and M. Karplus, "Evaluation of comparative protein modeling by MODELLER," *Proteins: Struct, Funct, Bioinf*, 23, pp. 318-326, 1995.

[73] "Discovery Studio Modeling Environment," in *Release 3.5*, ed. San Diego: Accelrys Software Inc., 2012, Accelrys Software Inc.

[74] M. -Y. Shen and A. Sali, "Statistical potential for assessment and prediction of protein structures," *Protein Sci*, 15, pp. 2507-2524, 2006.

[75] K. Tang, J. Zhang, and J. Liang, "Fast Protein Loop Sampling and Structure Prediction Using Distance-Guided Sequential Chain-Growth Monte Carlo Method," *PLoS Comput Biol*, 10, p. 1-16, 2014.

[76] L. Regad, J. Martin, G. Nuel, and A.-C. Camproux, "Mining protein loops using a structural alphabet and statistical exceptionality," *BMC Bioinformatics*, 11, p. 75, 2010.

[77] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J Appl Crystallogr*, 26, pp. 283-291, 1993.

[78] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, "Stereochemistry of polypeptide chain configurations," *J Mol Biol*, 7, pp. 95-99, 1963.

[79] S. Hovmoller, T. Zhou, and T. Ohlson, "Conformations of amino acids in proteins," *Acta Crystallogr, Sect D: Biol Crystallogr*, 58, pp. 768-776, 2002.

[80] V. K. Vyas, R. D. Ukawala, M. Ghate, and C. Chintha, "Homology modeling a fast tool for drug discovery: current perspectives," *Indian J Pharm Sci*, 74, pp. 1-17, 2012.

[81] L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey, and T. Schwede, "Protein structure homology modeling using SWISS-MODEL workspace," *Nature Protoc*, 4, pp. 1-13, 2008.

[82] C. J. J. Françoijs, J. P. G. Klomp, and R. M. A. Knegtel, "Sequence annotation of nuclear receptor ligand-binding domains by automated homology modeling," *Protein Eng, Des Sel*, 13, pp. 391-394, 2000.

[83] H.-J. Huang, H. W. Yu, C.-Y. Chen, C.-H. Hsu, H.-Y. Chen, K.-J. Lee, *et al.*, "Current developments of computer-aided drug design," *Journal of the Taiwan Institute of Chemical Engineers*, 41, pp. 623-635, 2010.

[84] S. Kalyaanamoorthy and Y.-P. P. Chen, "Structure-based drug design to augment hit discovery," *Drug Discovery Today*, 16, pp. 831-839, 2011.

[85] R. A. Laskowski, N. M. Luscombe, M. B. Swindells, and J. M. Thornton, "Protein clefts in molecular recognition and function," *Protein Sci*, 5, pp. 2438-2452, 1996.

[86] B. Nisius, F. Sha, and H. Gohlke, "Structure-based computational analysis of protein binding sites for function and druggability prediction," *J Biotechnol*, 159, pp. 123-134, 2012.

[87] D. Ghersi and R. Sanchez, "Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures," *J Struct Funct Genomic*, 12, pp. 109-117, 2011.

[88] S. Henrich, O. M. Salo-Ahen, B. Huang, F. F. Rippmann, G. Cruciani, and R. C. Wade, "Computational approaches to identifying and characterizing protein binding sites for ligand design," *J Mol Recognit*, 23, pp. 209-219, 2010.

[89] Z. Zhang, Y. Li, B. Lin, M. Schroeder, and B. Huang, "Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction," *Bioinformatics,* 27, pp. 2083-2088, 2011.

[90] J. Yu, Y. Zhou, I. Tanaka, and M. Yao, "Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere," *Bioinformatics,* 26, pp. 46-52, 2010.

[91] P. J. Goodford, "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules," *J Med Chem,* 28, pp. 849-857, 1985.

[92] A. T. Laurie and R. M. Jackson, "Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites," *Bioinformatics,* 21, pp. 1908-16, 2005.

[93] Z. Wang, L. Sun, H. Yu, Y. Zhang, W. Gong, H. Jin, *et al.*, "Binding Mode Prediction of Evodiamine within Vanilloid Receptor TRPV1," *Int J Mol Sci,* 13, pp. 8958-69, 2012.

[94] C. M. Venkatachalam, X. Jiang, T. Oldfield, and M. Waldman, "LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites," *J Mol Graphics Modell,* 21, pp. 289-307, 2003.

[95] D. G. Levitt and L. J. Banaszak, "POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids," *J Mol Graph,* 10, pp. 229-34, 1992.

[96] M. Hendlich, F. Rippmann, and G. Barnickel, "LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins," *J Mol Graphics Modell,* 15, pp. 359-63, 389, 1997.

[97] H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining docking and molecular dynamic simulations in drug design," *Med Res Rev,* 26, pp. 531-68, 2006.

[98] C. A. Lipinski, "Drug-like properties and the causes of poor solubility and poor permeability," *J Pharmacol Toxicol Methods,* 44, pp. 235-49, 2000.

[99] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv Drug Delivery Rev,* 23, pp. 3-25, 1997.

[100] L. K. Gavrin and E. Saiah, "Approaches to discover non-ATP site kinase inhibitors," *Med Chem Comm,* 4, pp. 41-51, 2013.

[101] D. Huang, T. Zhou, K. Lafleur, C. Nevado, and A. Caflisch, "Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis," *Bioinformatics,* 26, pp. 198-204, 2010.

[102] P. A. Schwartz and B. W. Murray, "Protein kinase biochemistry and drug discovery," *Bioorg Chem,* 39, pp. 192-210, 2011.

[103] A. Vulpetti and R. Bosotti, "Sequence and structural analysis of kinase ATP pocket residues," *Farmaco,* 59, pp. 759-65, 2004.

[104] E. A. Martis, R. Radhakrishnan, and R. R. Badve, "High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview," *Journal of Applied Pharmaceutical Science,* 01, pp. 02-10, 2011.

[105] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, *et al.*, "Impact of high-throughput screening in biomedical research," *Nat Rev Drug Discovery,* 10, pp. 188-195, 2011.

[106] R. E. Hubbard, "Structure-based drug discovery and protein targets in the CNS," *Neuropharmacology,* 60, pp. 7-23, 2011.

[107] E. Yuriev, M. Agostino, and P. A. Ramsland, "Challenges and advances in computational docking: 2009 in review," *J Mol Recognit,* 24, pp. 149-164, 2011.

[108] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *J Mol Biol,* 267, pp. 727-48, 1997.

[109] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, *et al.*, "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility," *J Comput Chem,* 30, pp. 2785-91, 2009.

[110] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *J Mol Biol,* 261, pp. 470-489, 1996.

[111] S. N. Rao, M. S. Head, A. Kulkarni, and J. M. LaLonde, "Validation studies of the site-directed docking program LibDock," *J Chem Inf Model,* 47, pp. 2159-71, 2007.

[112] G. Wu, D. H. Robertson, C. L. Brooks, and M. Vieth, "Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm," *J Comput Chem,* 24, pp. 1549-1562, 2003.

[113] M. Taufer, M. Crowley, D. J. Price, A. A. Chien, and C. L. Brooks, "Study of a highly accurate and fast protein–ligand docking method based on molecular dynamics," *Concurrency and Computation: Practice and Experience,* 17, pp. 1627-1641, 2005.

[114] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications," *Nat Rev Drug Discovery,* 3, pp. 935-949, 2004.

[115] M. Xue, M. Zheng, B. Xiong, Y. Li, H. Jiang, and J. Shen, "Knowledge-based scoring functions in drug design. 1. Developing a target-specific method for kinase-ligand interactions," *J Chem Inf Model,* 50, pp. 1378-86, 2010.

[116] F. Sousa, P. A. Fernandes, and M. Joa, "Protein–Ligand Docking: Current Status and Future Challenges " *Proteins: Struct, Funct, Bioinf,* 65, pp. 15–26, 2006.

[117] A. Krammer, P. D. Kirchhoff, X. Jiang, C. M. Venkatachalam, and M. Waldman, "LigScore: a novel scoring function for predicting binding affinities," *J Mol Graphics Modell,* 23, pp. 395–407, 2005.

[118] I. Muegge, "PMF scoring revisited," *J Med Chem,* 49, pp. 5895-902, 2006.

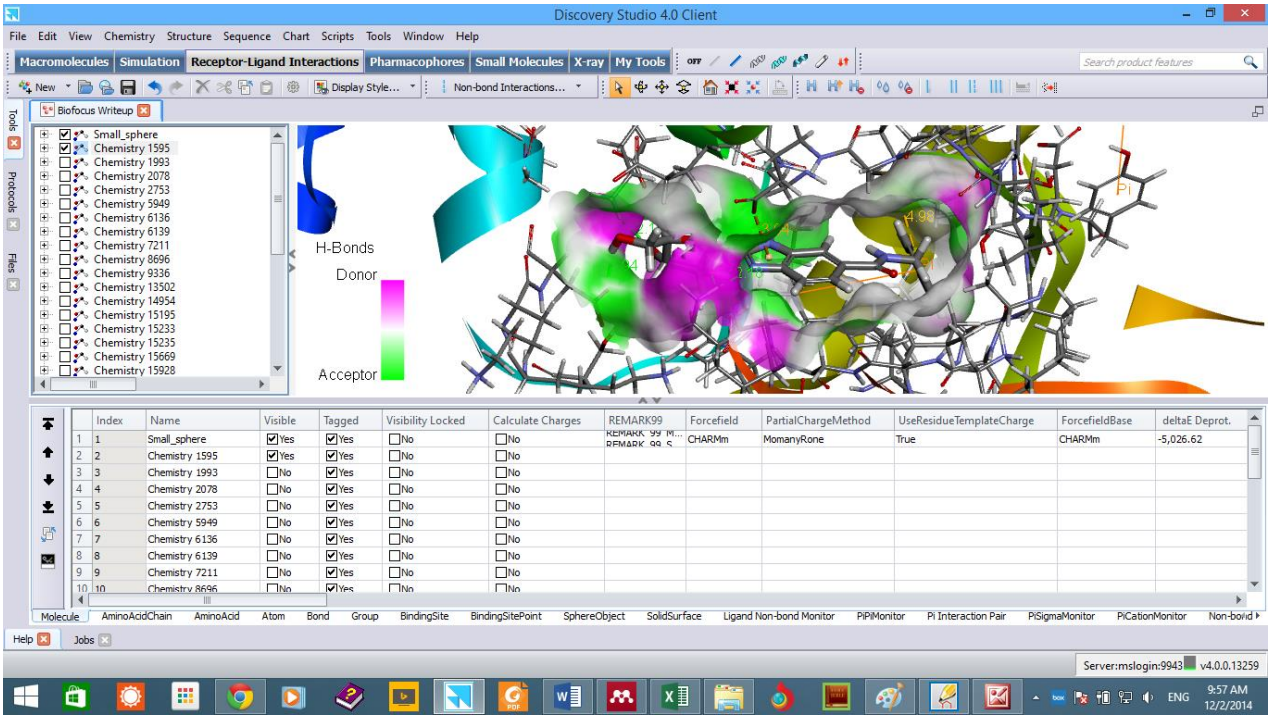
[119] H.-Y. Zhao, H. Wei, and X. Wang, "The Reciprocal Interaction of Small Molecule Protein Kinase Inhibitors and ATP-Binding Cassette Transporters in Targrted Cancer Therapy," *J Cancer Res Updates,* 2, pp. 68-86, 2013.

[120] G. Francois, C. M. Passreiter, H. J. Woerdenbag, and M. Van Looveren, "Antiplasmodial activities and cytotoxic effects of aqueous extracts and sesquiterpene lactones from Neurolaena lobata," *Planta Med,* 62, pp. 126-9, 1996.

[121] R. W. Rickards, J. M. Rothschild, A. C. Willis, N. M. de Chazal, J. Kirk, K. Kirk, *et al.*, "Calothrixins A and B, novel pentacyclic metabolites from Calothrix cyanobacteria with potent activity against malaria parasites and human cancer cells," *Tetrahedron,* 55, pp. 13513-13520, 1999.

Appendices

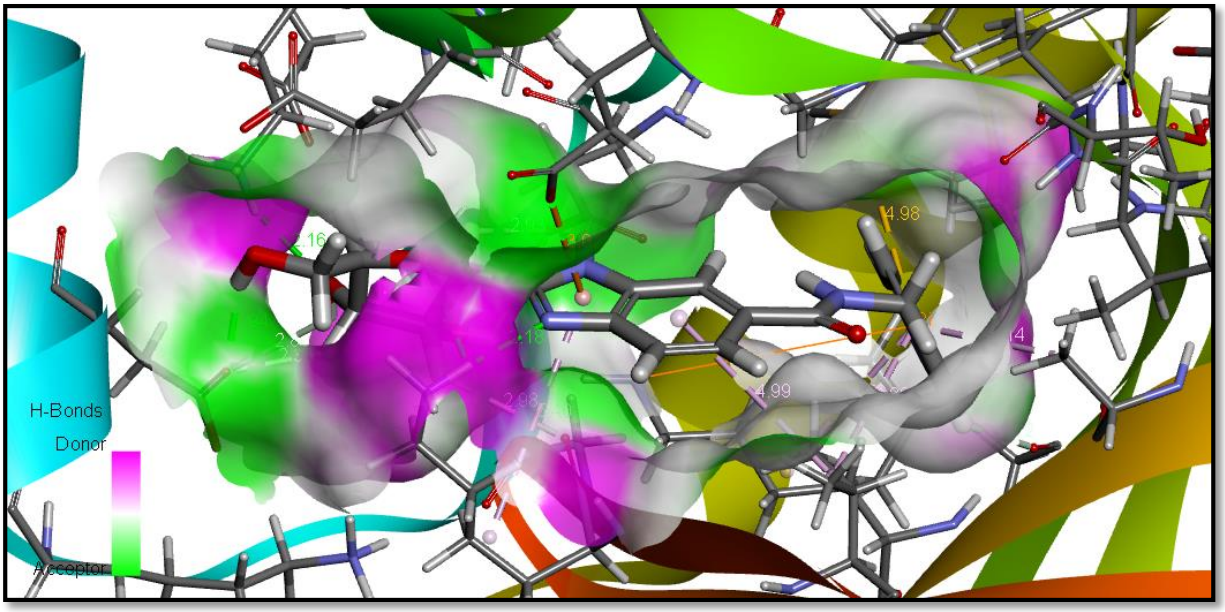
Appendix A



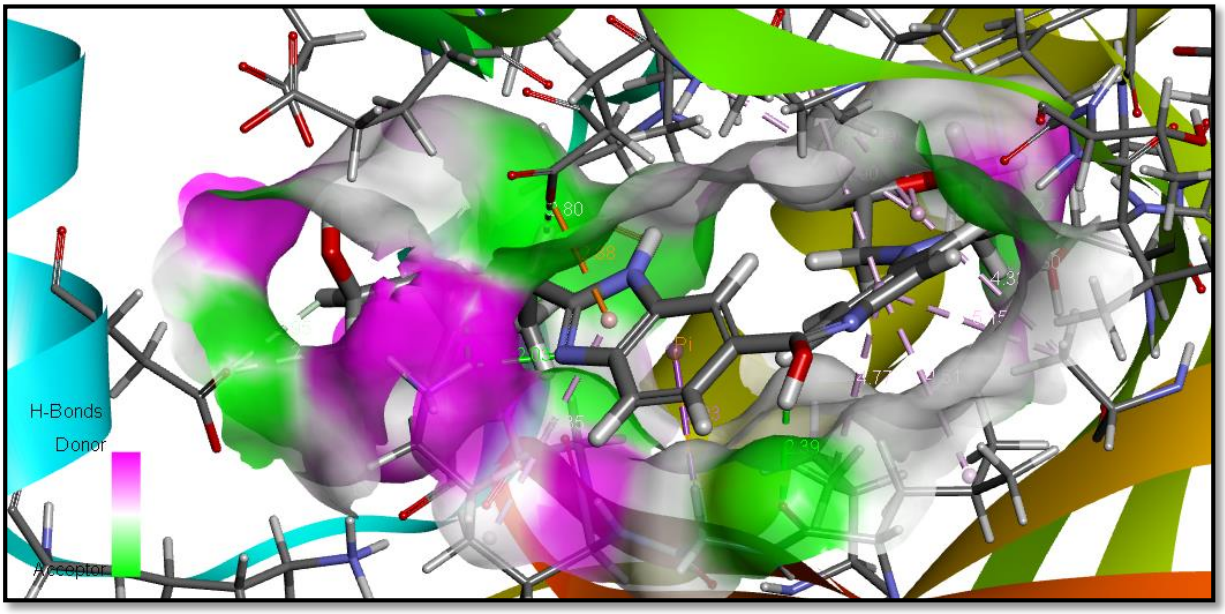
Discovery Studio software used for all computational studies for this research project. The license for the software was accessed via Centre for High Performance Computing (CHPC).

Appendix B

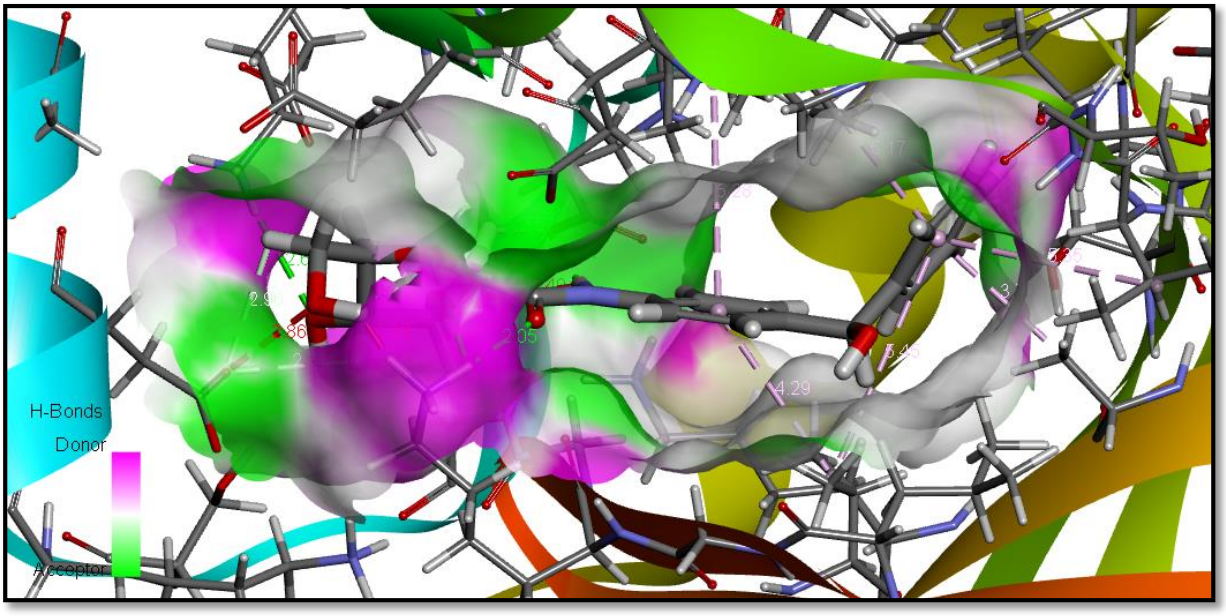
Conformations of prioritized compounds.



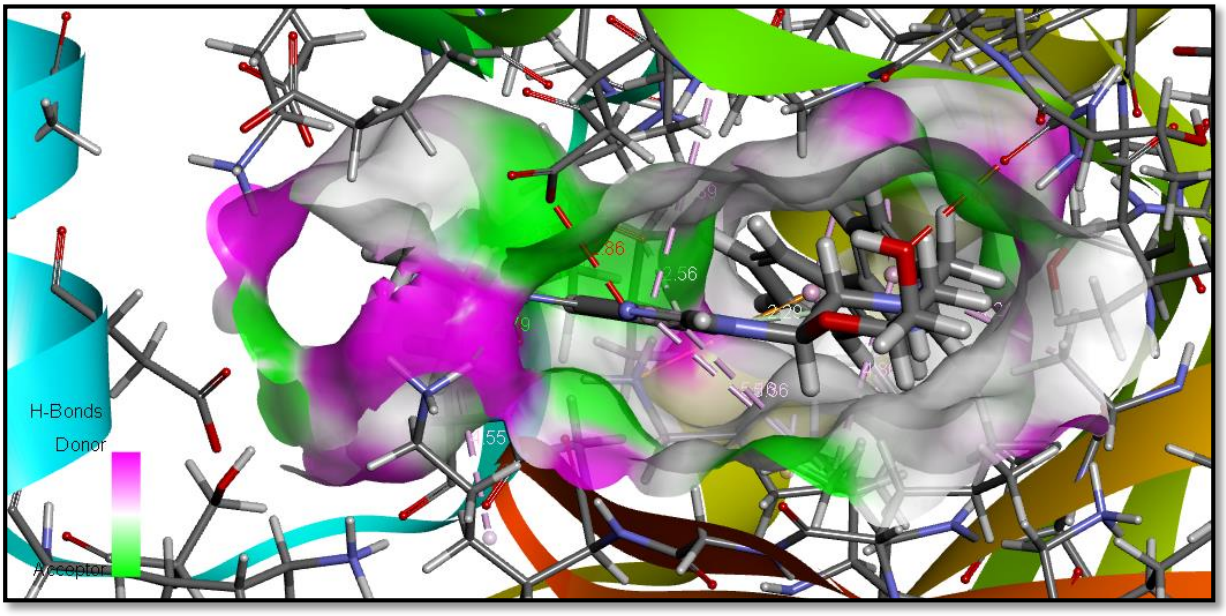
TM1595



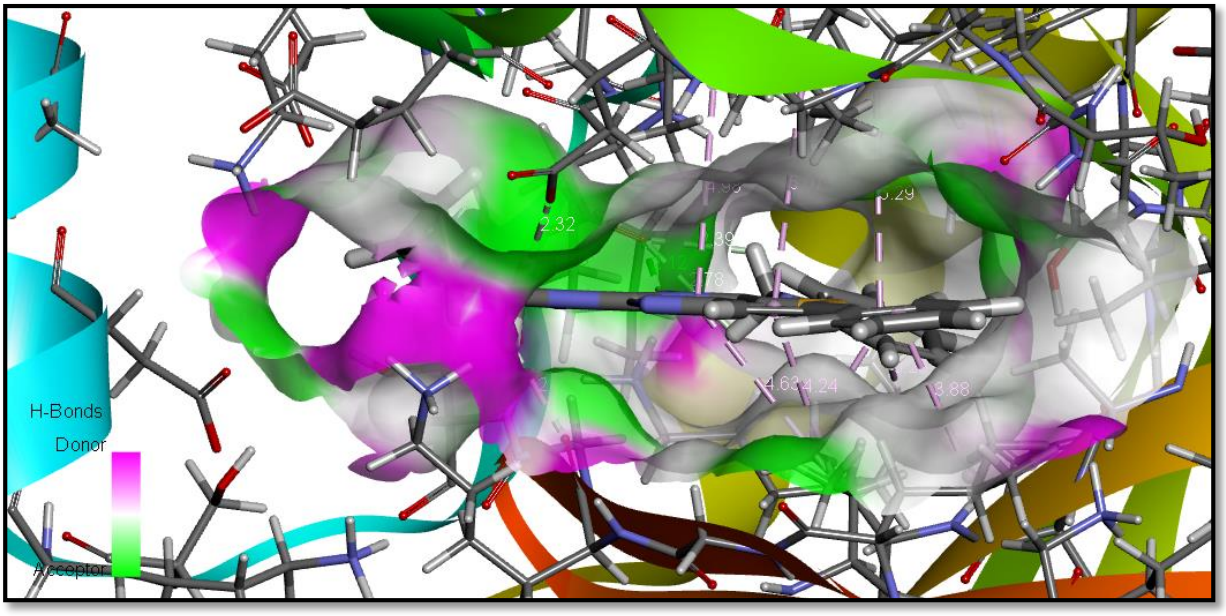
TM1993



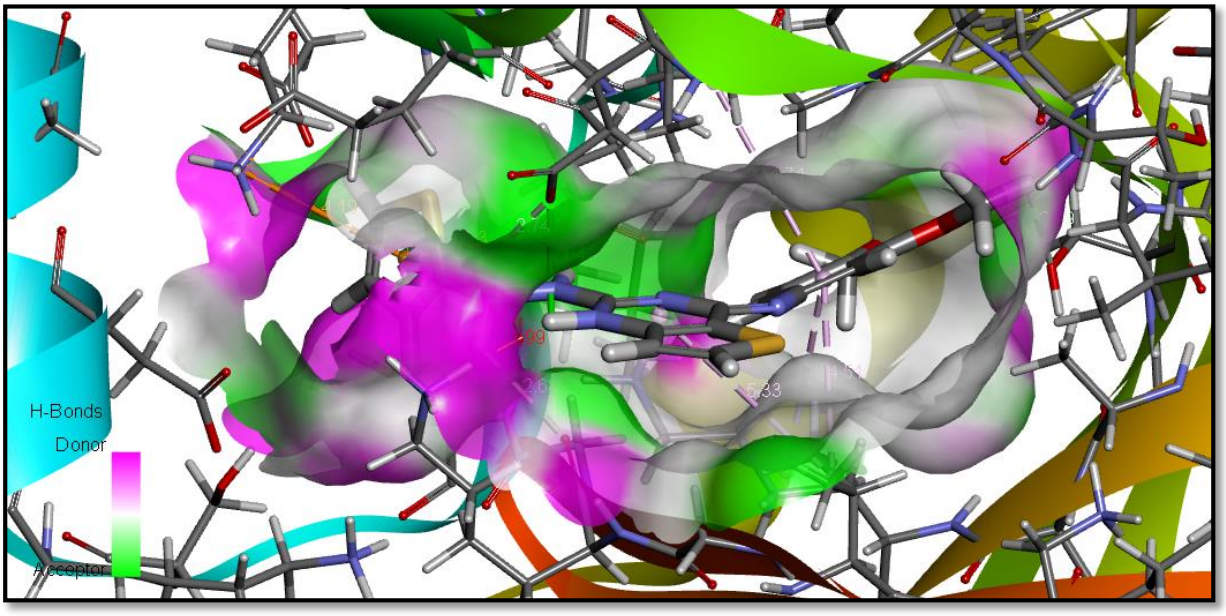
TM2078



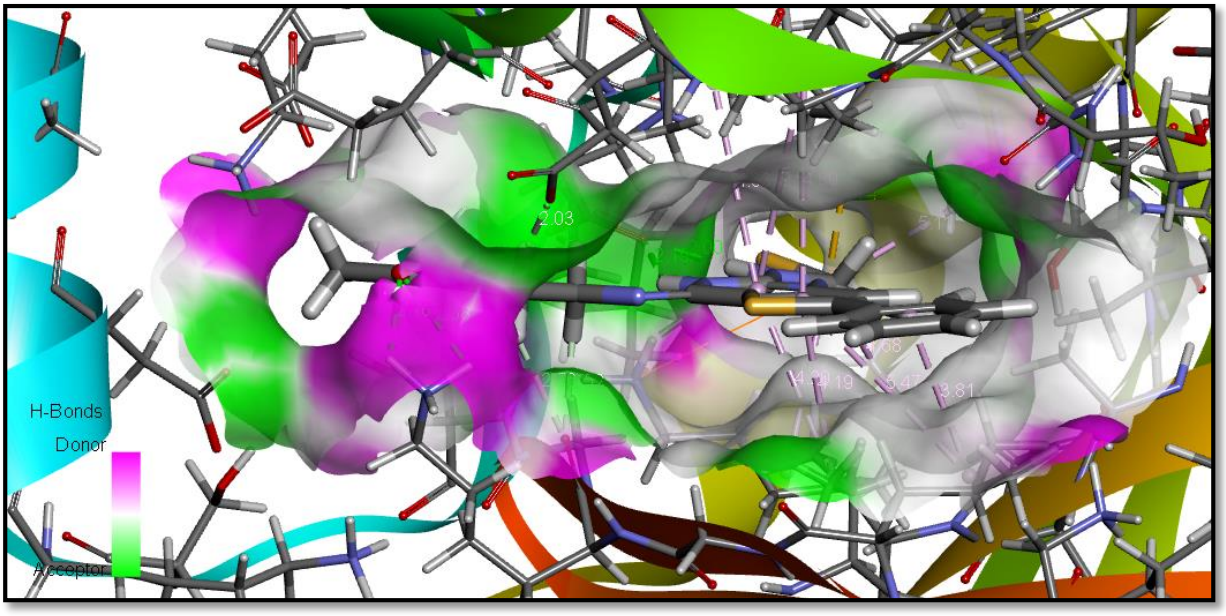
TM2753



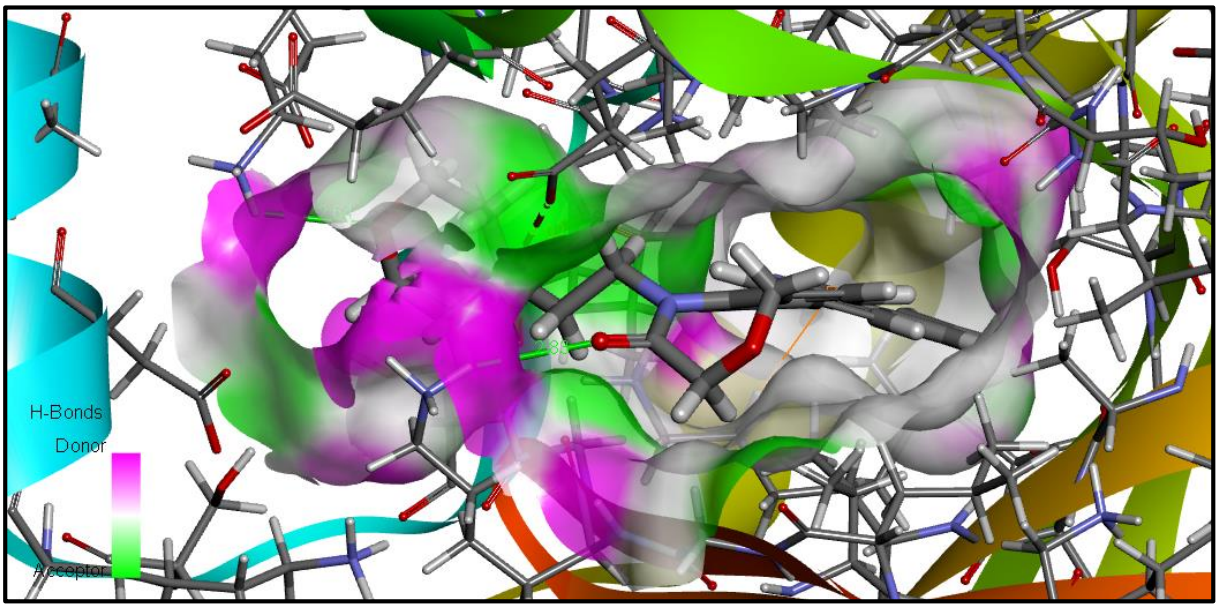
TM5949



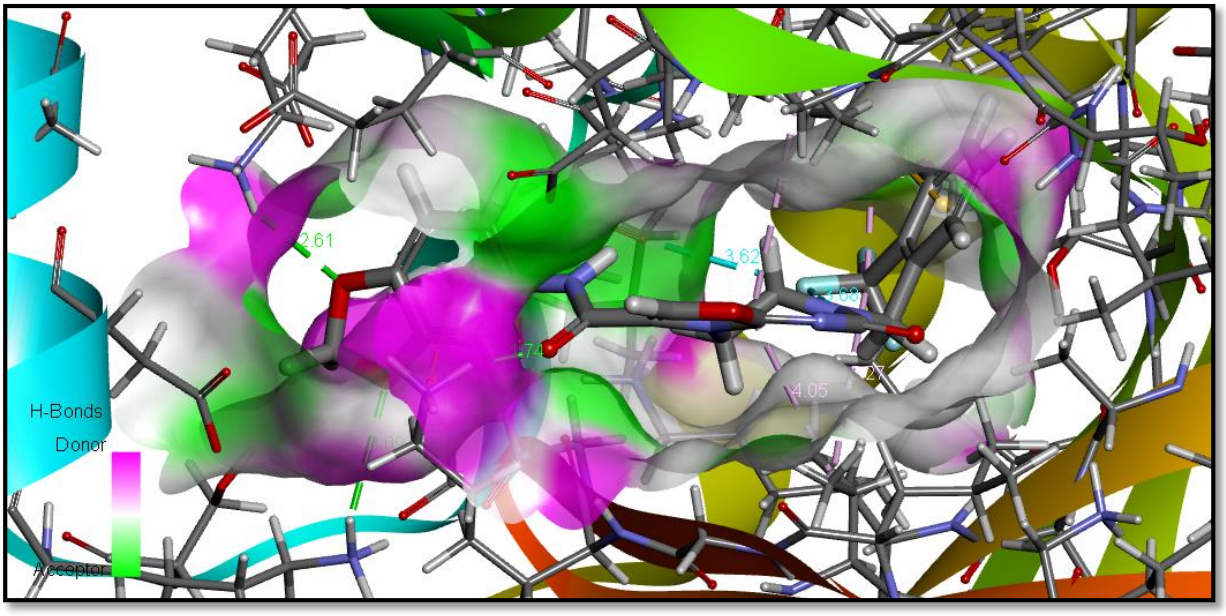
TM6136



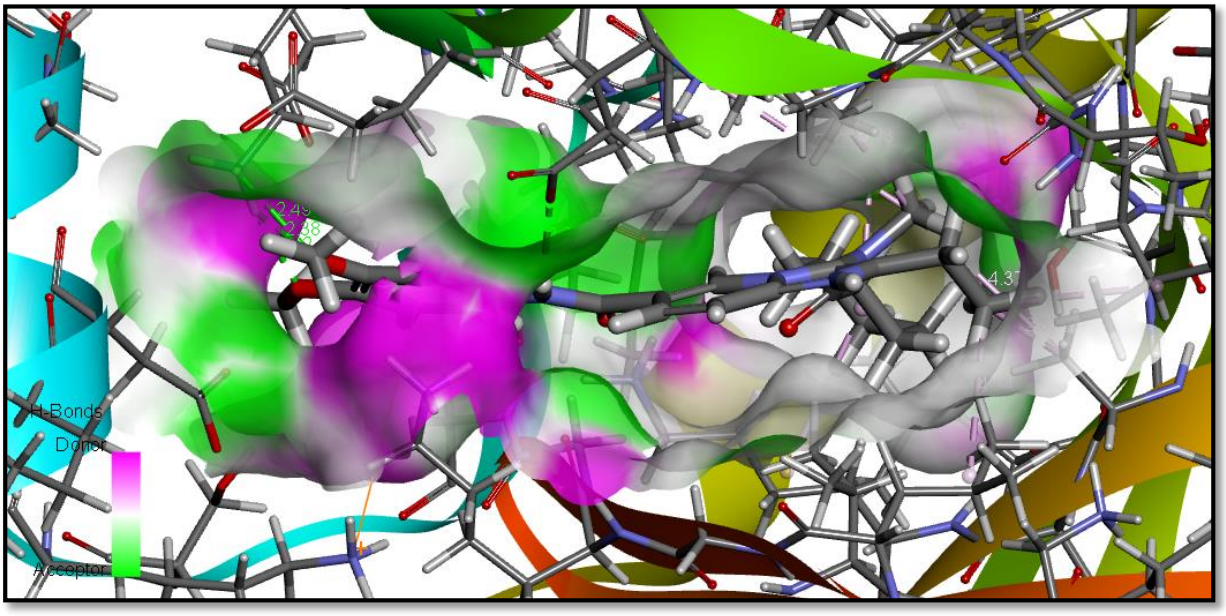
TM6139



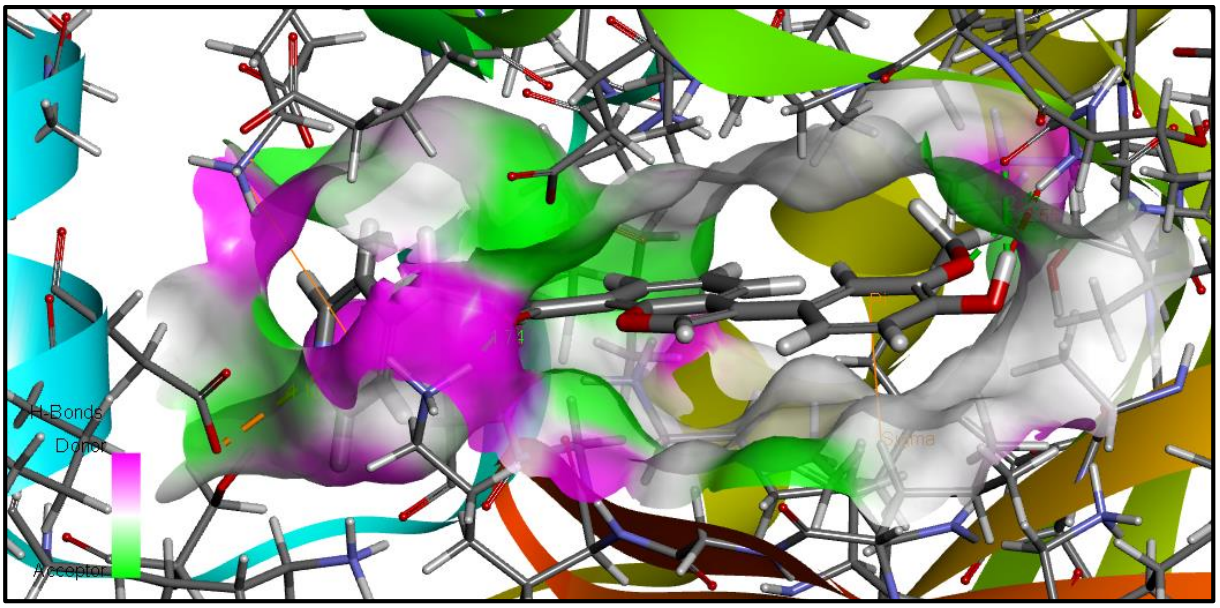
TM7211



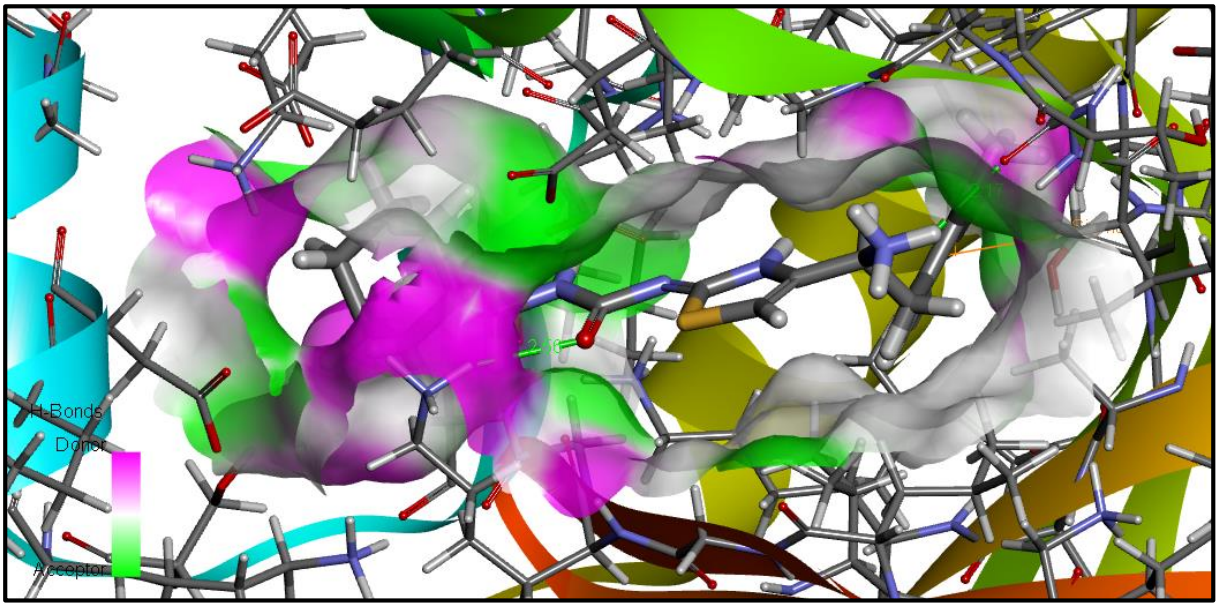
TM8696



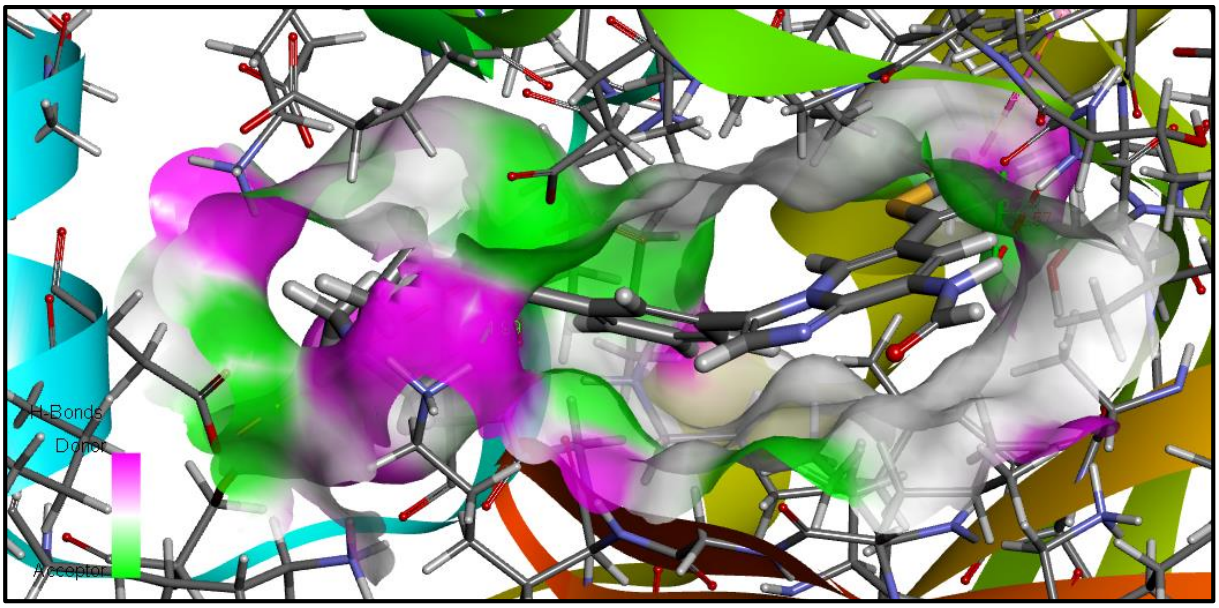
TM9336



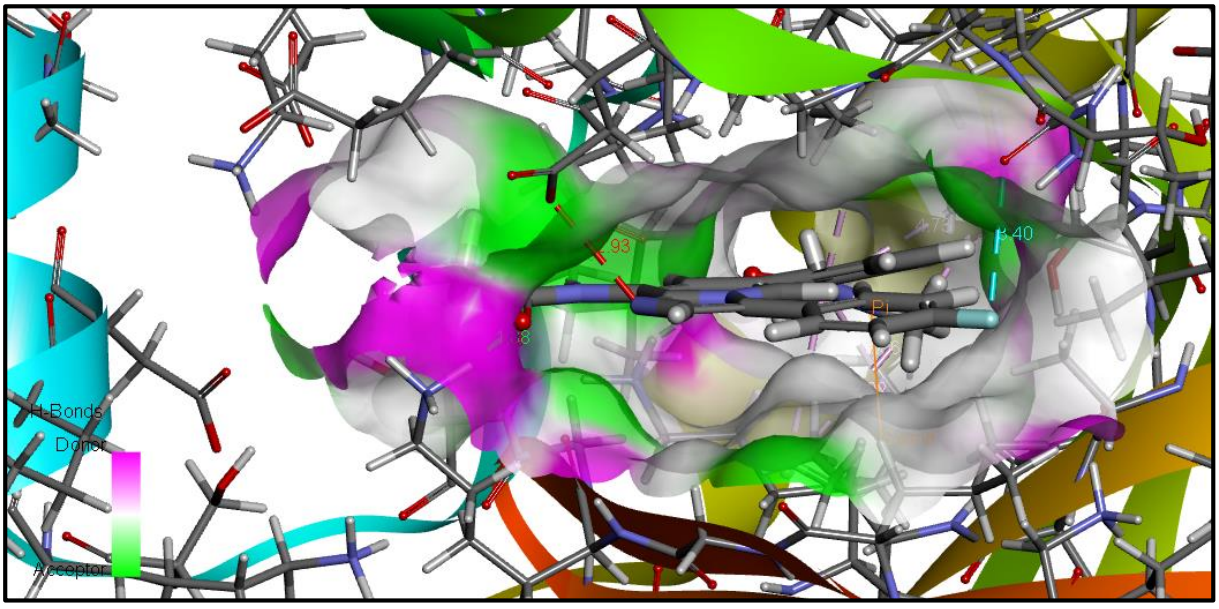
TM13502



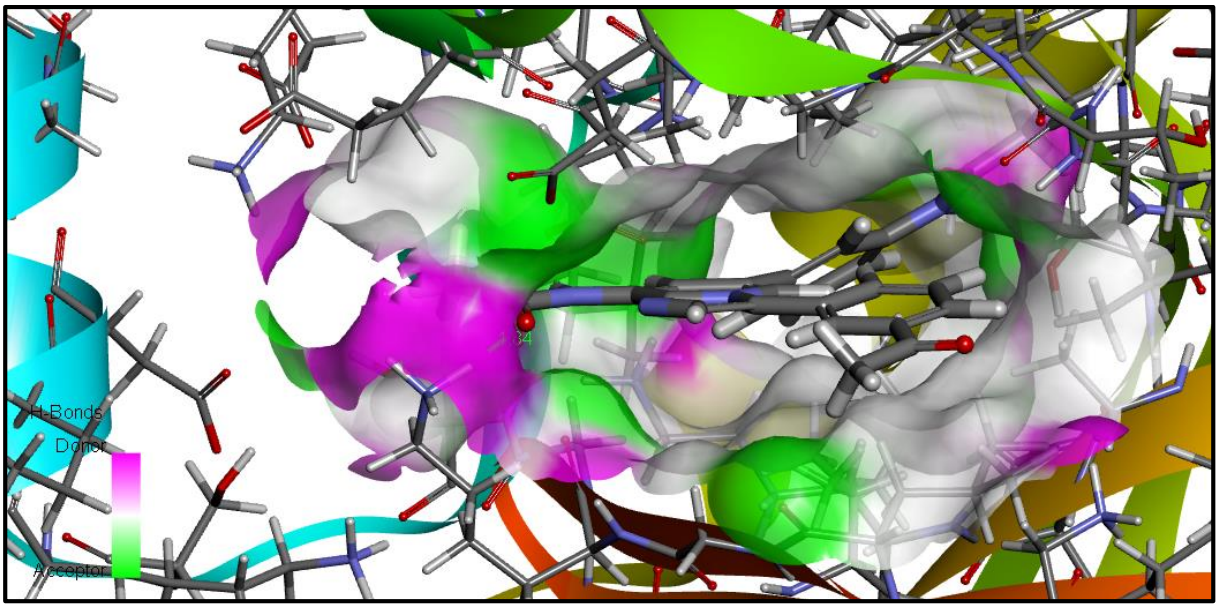
TM14954



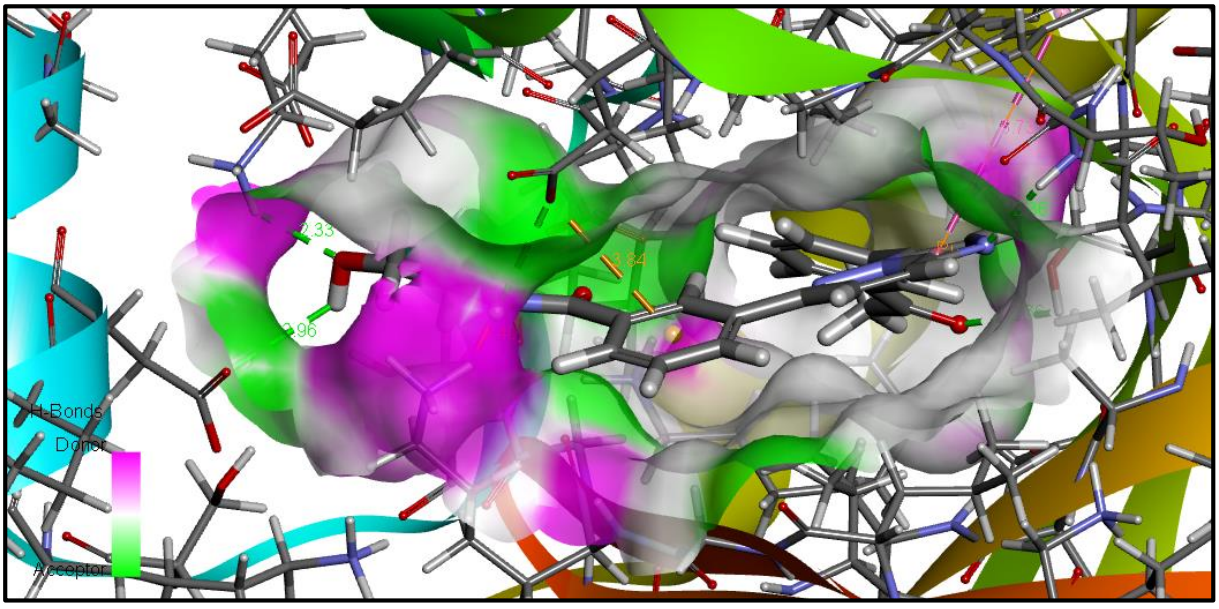
TM15195



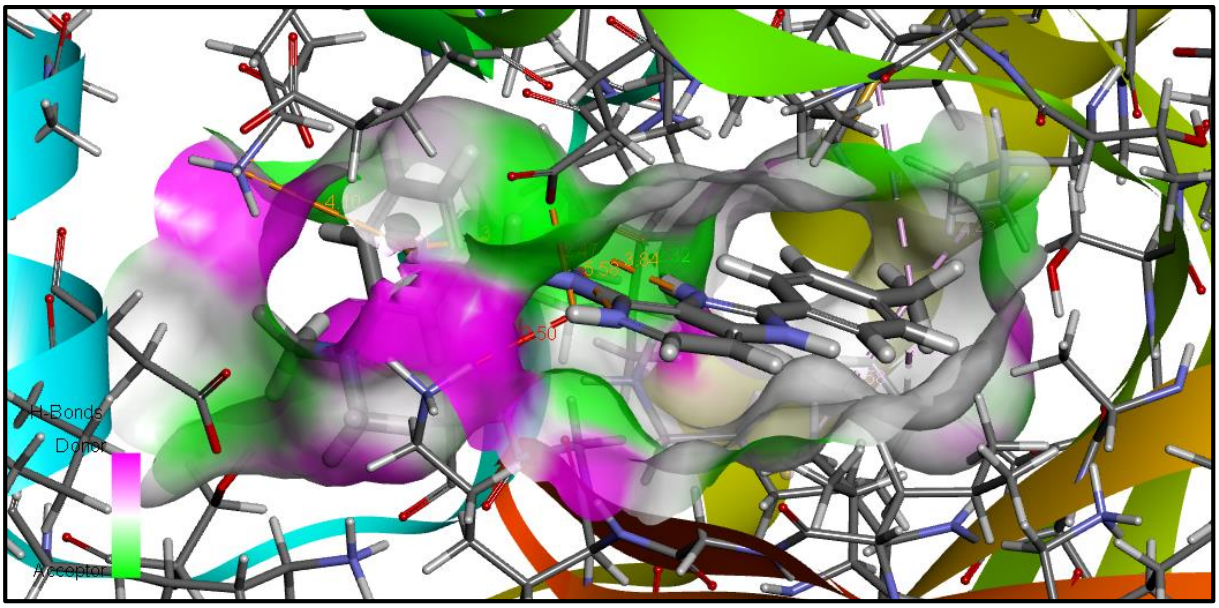
TM15233



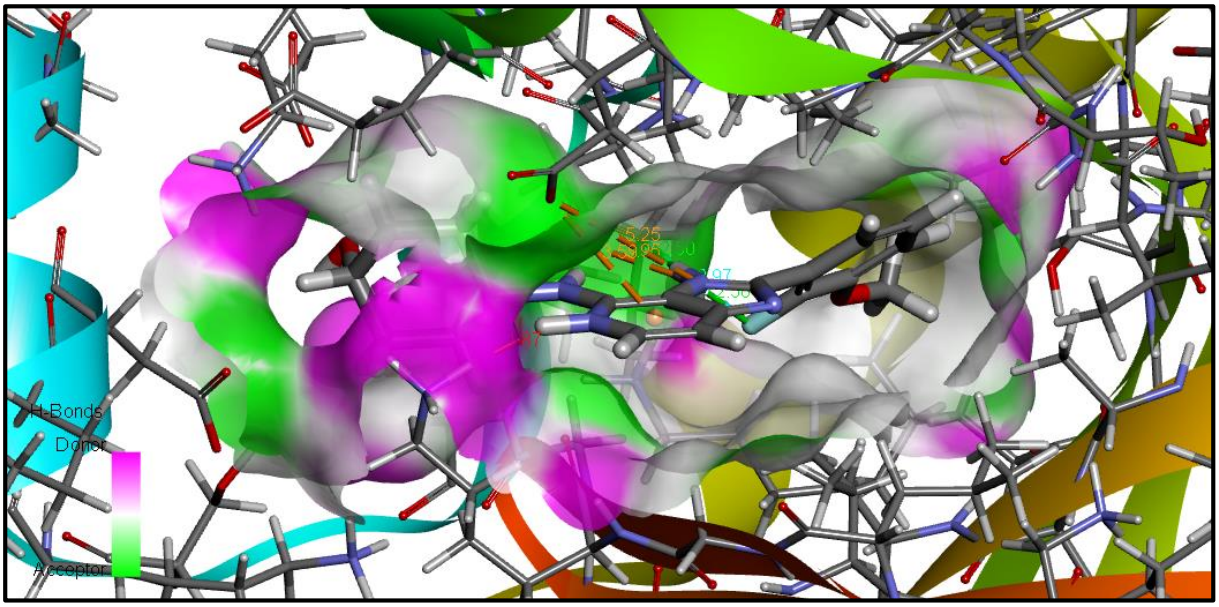
TM15235



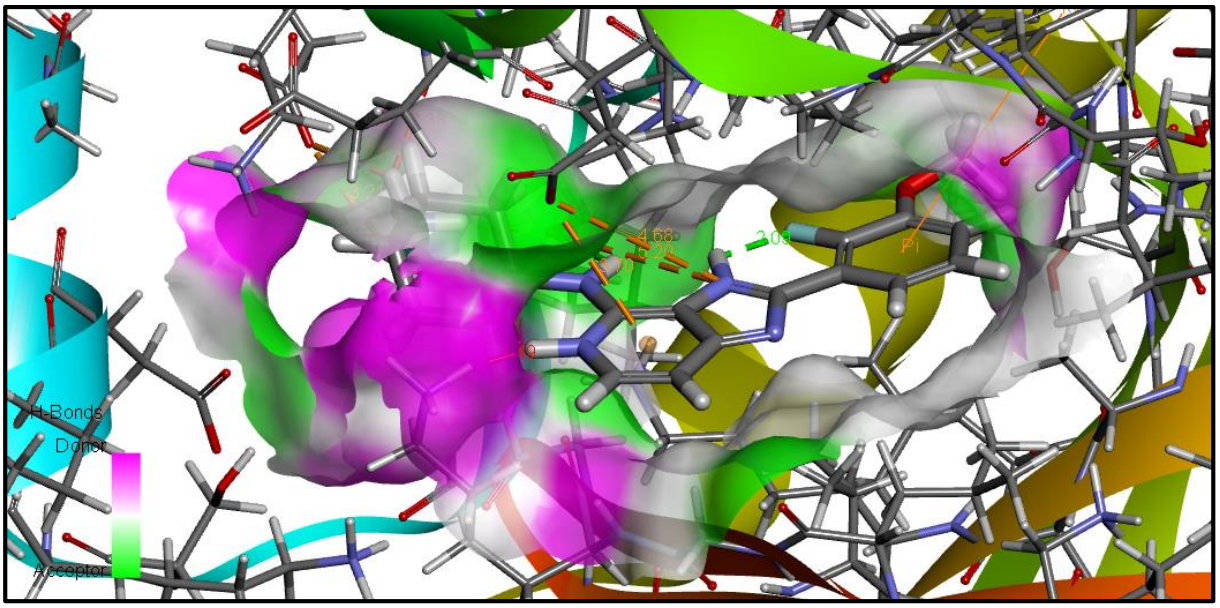
TM15669



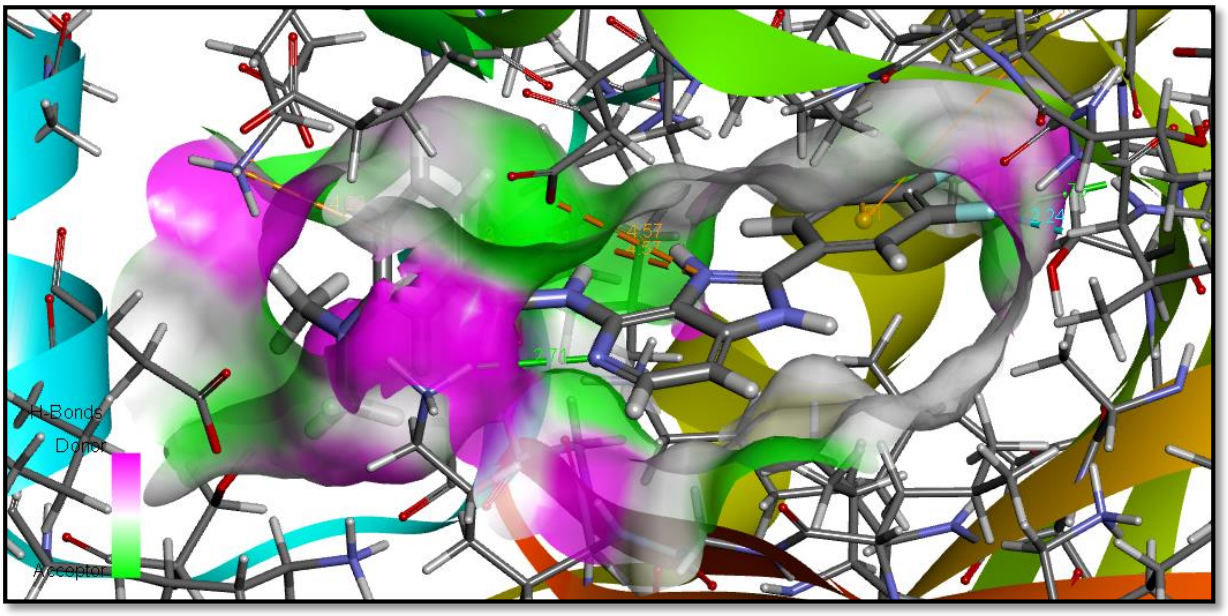
TM15928



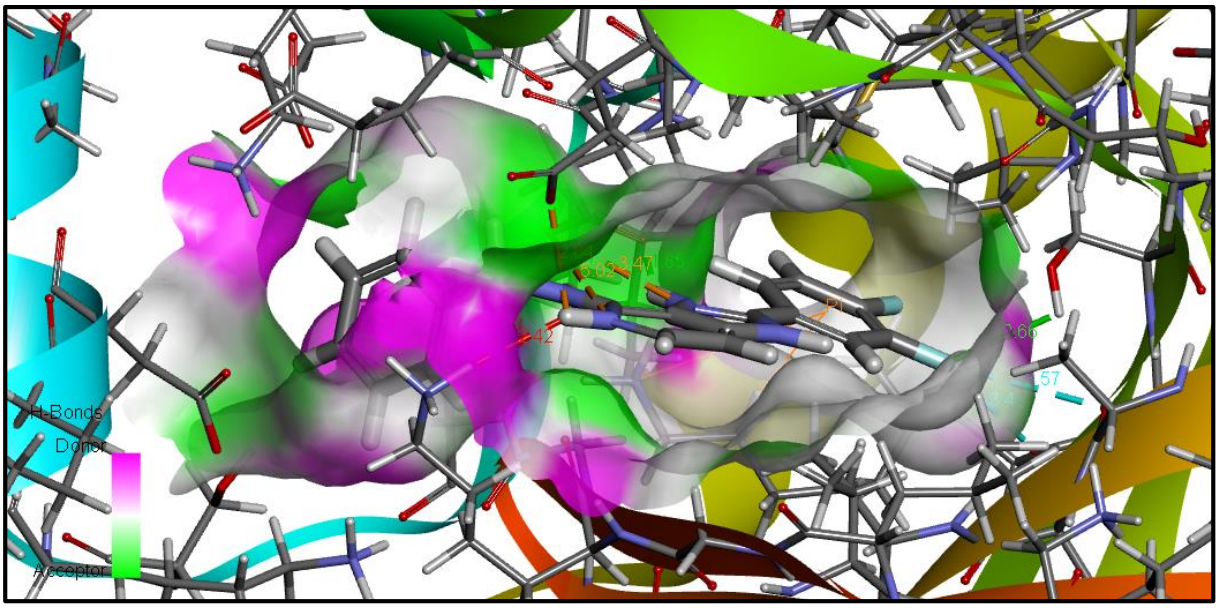
TM15930



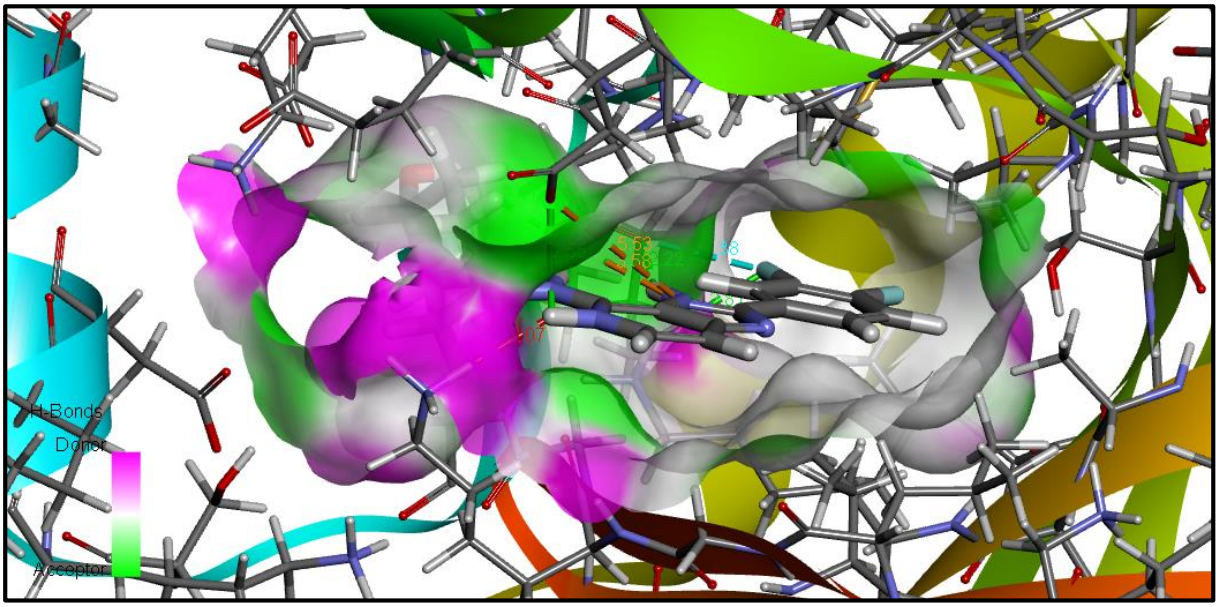
TM15940



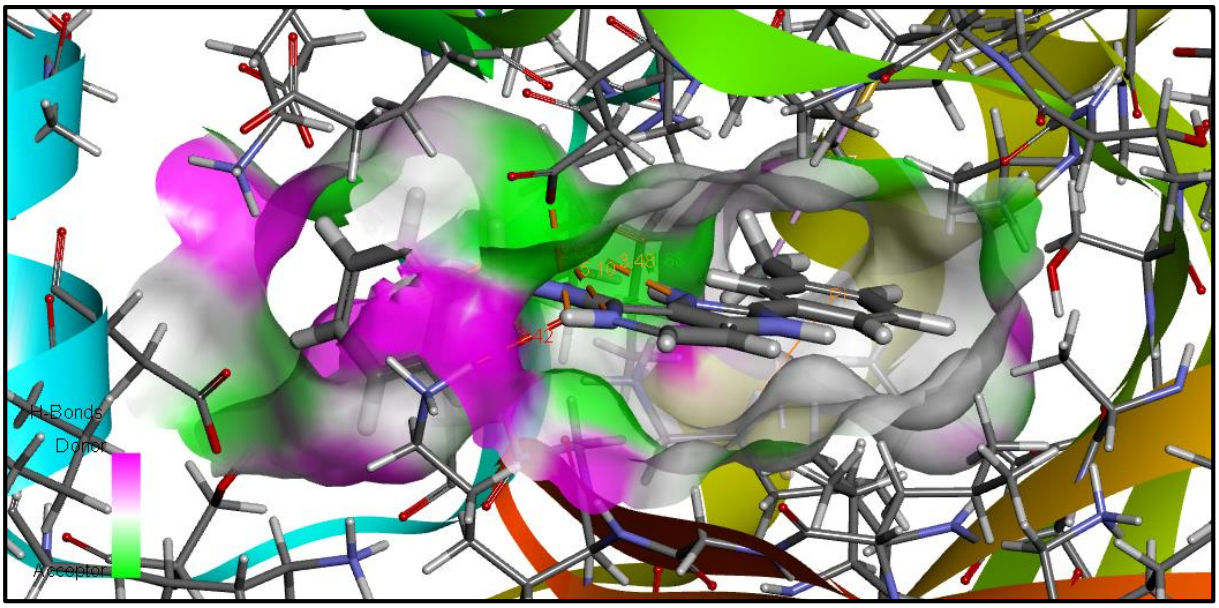
TM16028



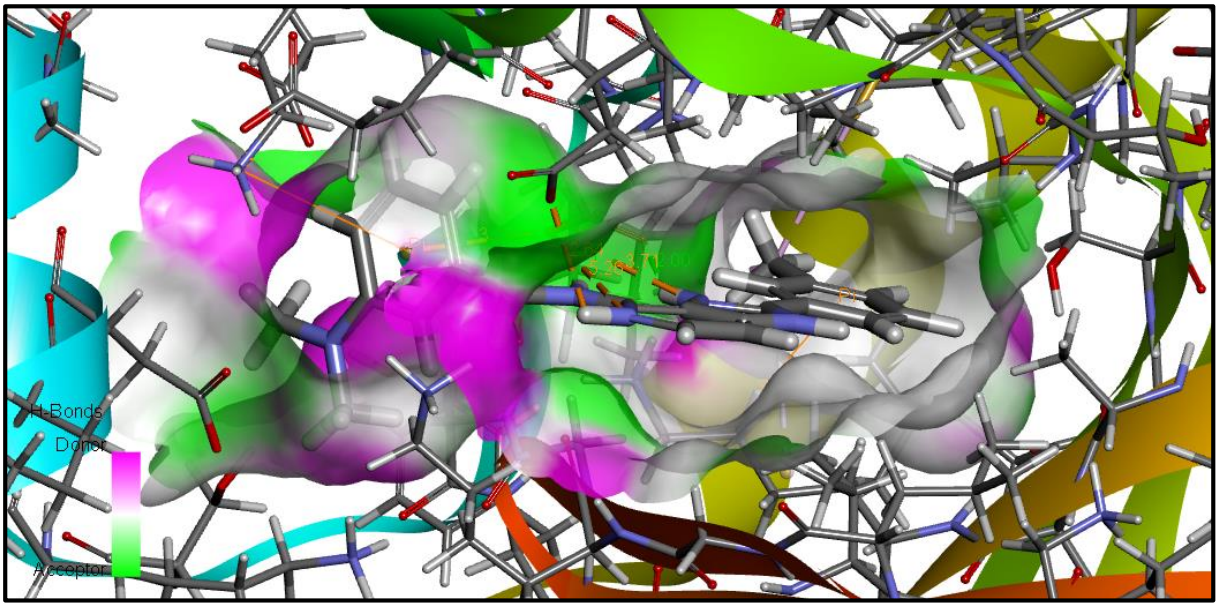
TM16037



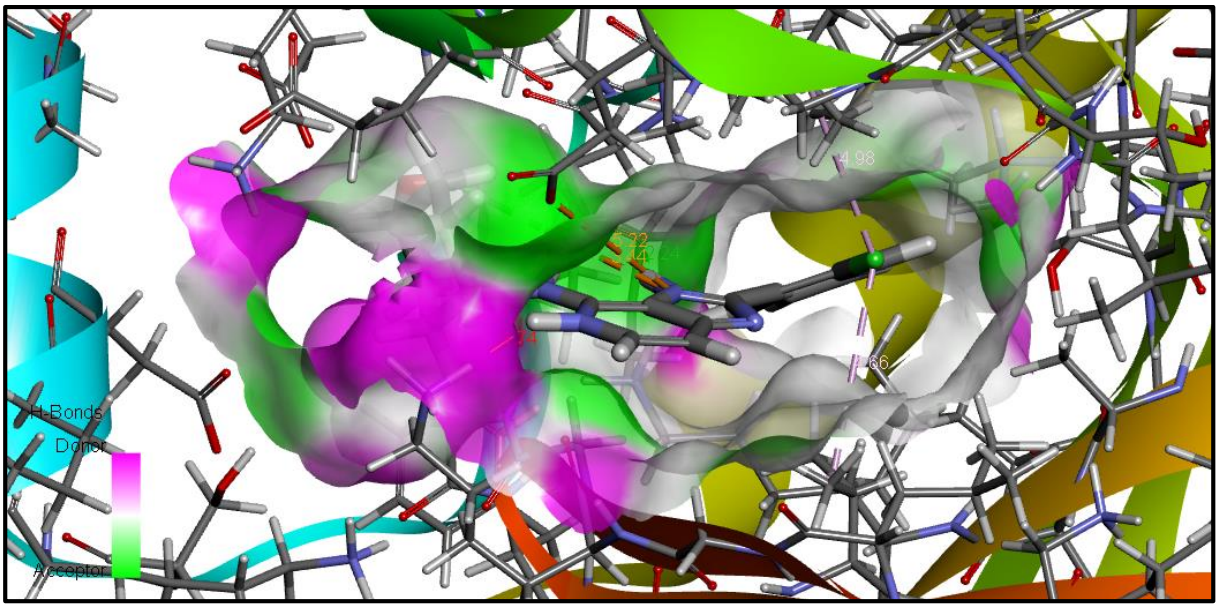
TM16047



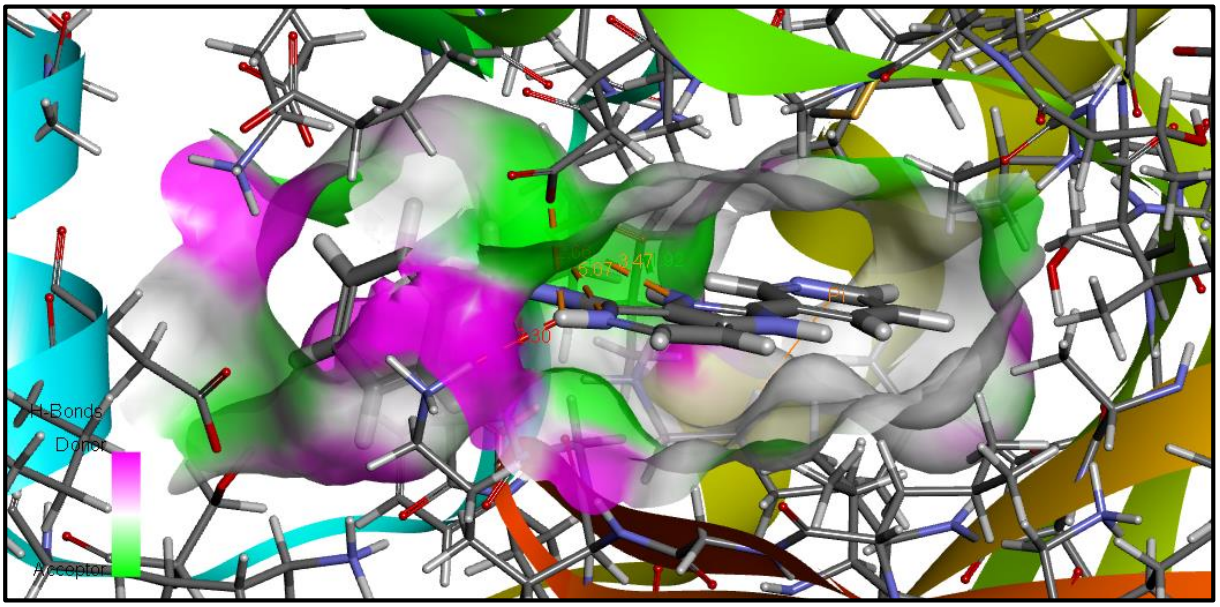
TM16054



TM16065



TM16098



TM19688