

# *In Silico* Evaluation of Inhibitors of *Plasmodium Falciparum* AP2-I Transcription Factor and *Plasmodium Falciparum* Bromodomain Protein 1

David Oladejo (✉ [oladejo.oladoke@stu.cu.edu.ng](mailto:oladejo.oladoke@stu.cu.edu.ng))

Covenant University

Gbolahan Oduselu

Covenant University

Titilope Dokunmu

Covenant University

Itunuoluwa Isewon

Covenant University

Esther Okafor

Covenant University

Emeka E.J. Iweala

Covenant University

Ezekiel F. Adebisi

Covenant University

---

## Research Article

### Keywords:

**Posted Date:** February 4th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1318124/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

Recent treatment failures observed for Artemisinin-based combination therapy (ACT) have raised concerns about the efficacy of the front-line drug to treat malaria and the need to develop a new antimalarial drug regimen. *Plasmodium falciparum* Apicomplexan Apetala 2 Invasion (*PfAP2-I*) transcription factor (TF) is a protein that regulates the expression of a subset of gene families involved in *Plasmodium falciparum* red blood cell (RBC) invasion. *PfAP2-I* associates with several chromatin proteins, including the *Plasmodium falciparum* bromodomain protein 1 (*PfBDP1*) and that complex formation is associated with transcriptional regulation. Inhibiting *PfAP2-I* TF and *PfBDP1* with small molecules represents a potential new antimalarial therapeutic target to combat drug resistance, which this study aims to achieve.

## Methods

The 3D model structure of *PfAP2-I* was predicted *ab initio* using ITASSER and ROBETTA prediction tools and was validated using Errat and Procheck from the Save server 6.0. The crystal structure of *PfBDP1* was also retrieved from Protein Data Bank (PDB) ([www.rcsb.org/structure/7M97](http://www.rcsb.org/structure/7M97)) and Computed Atlas of Surface Topography of proteins (CASTp) 3.0 and *ConCavity* were used to predict the active sites of the *PfAP2-I* and *PfBDP1* 3D7 structures. Pharmacophore modeling of the control ligand (3W7 from COACH server) and modeled 3D structure of *PfAP2-I* was carried out using the Pharmit server to obtain several compounds for docking analysis. Chimera software was used to remove the complexed ligands, and the modeled protein structure was defined as a receptor. Virtual screening and post-screening studies were conducted using AutoDock vina and LigPlot, respectively. The designed ligands' toxicity predictions and *in silico* drug-likeness were performed using the Swiss ADME predictor and OSIRIS Property Explorer.

## Results

The result of the modeled protein from the ROBETTA prediction tool was prioritized based on structure validation results of 96.827 for ERRAT and 90.2% of the amino acid residues in the most favored region for the Ramachandran plot. A total of 8656 compounds obtained from nine (9) databases on the Pharmit server were used to prepare the ligand library and screened against the prepared 3D model structure of *PfAP2-I*, considering the active sites predicted from CASTp and *ConCavity*. Six (6) best hits were selected based on the binding affinity of the ligands to the active site *PfAP2-I* and were considered for post-screening analyses. The six best hits exhibited dock scores between -9.9 and -10.2 kcal/mol for *PfAP2-I* and between -8.5 and -9.4 kcal/mol for *PfBDP1*. The best hits also had lower binding energies in the *PfAP2-I* docking model when compared to the reference compound, ChEMBL3359262 (-8.4 kcal/mol) and the standard drug, chloroquine (-4.2 kcal/mol). In the *PfBDP1* docking model, the reference compound, ChEMBL3359262 and the standard drug, chloroquine has binding affinities of -8.3 and -6.1 kcal/mol respectively. The compounds with the best dock scores are ZINC97139187 (-10.2 kcal/mol) for *PfAP2-I* and MCULE-6567089130 (-9.4 kcal/mol) for *PfBDP1*. For the ADMET properties, compound ZINC97139187 had the highest drug score of 0.63, followed by compound 154861216, MCULE-6567089130 and 57405339 with drug scores of 0.58, 0.47 and 0.47 respectively (all higher than that of the standard drug - chloroquine of 0.25).

## Conclusions

The good, estimated binding energies and drug scores observed for compound ZINC97139187 and compound MCULE-6567089130 suggest that they can be considered possible *PfAP2-I* and *PfBDP1* inhibitors. Further pre-clinical experimental validations should be carried out to ascertain the efficacy of these predicted best hits.

## Background

Many antimalarial drugs have been produced over the years, but resistant parasite strains have formed against most of them, including chloroquine, pyrimethamine, and proguanil [1]. Recent treatment failures with artemisinin-based combination therapy (ACT) have raised concerns about the loss of the highly-effective treatment currently available to treat malaria [2]. The licensed antimalarial drug's poor efficacy, combined with the spread of antimalarial drug resistance, necessitates the development of an innovative strategy to identify novel antimalarial compounds [3]. Despite tremendous advances in understanding malaria epidemiology and the availability of several therapeutic options, malaria remains one of the leading global causes of death, with children accounting for a large proportion of those affected [4]

*In silico* techniques have been successful and have become powerful tools in the search to cure disease [5] and reducing the use of animal models in pharmacological research, assisting in the rational design of novel and safe drug candidates, and repositioning marketed drugs [6]. They are vital in identifying viable therapeutic candidates at a low cost and time by utilizing sophisticated computers and information technology to speed up drug discovery, lead optimization, drug development, and design [7]. *In silico* methods such as molecular dynamics simulations, molecular docking, drugs-likeness prediction, ADMET (absorption, distribution, metabolism, elimination, and toxicity) studies are used to screen candidate drugs/molecules from various databases/libraries [7]. These methods have been proposed to recognize and select therapeutic relevant targets, study the molecular basis of drug-receptor complexes interactions, structurally characterize ligand binding sites on biological targets, design de novo target-specific compound libraries, predict target protein structure, identify hit compound by ligand- and structure-based virtual screening, estimate binding free energy between a ligand and receptor, and optimize high-affinity ligands [8].

The idea of targeting transcription and transcription factors (TFs) for drug therapy was long considered a "Sisyphian task" but recent work in drug discovery has shown the direct modulation of transcription factor function by small molecules [9–11]. Transcription factors (TFs) are proteins that bind to DNA sequences and control the stream of genetic information from DNA to mRNA [12]. TFs, along with other proteins in a complex, control *Plasmodium falciparum* gene expression by promoting (activator) or blocking (repressor) the recruitment of RNA polymerase to specific genes during the intra-erythrocytic

development cycle (IDC) in the red blood cells (RBC) [13, 14]. In antimalarial drug design, TFs as drug targets have enormous potential to be drug resistance-free because targeting TFs affects many genes instead of one gene for enzymatic sites-based drugs [11]. One of such important TFs found across the different *Plasmodium spp* is the Apetala 2 – Invasion (AP2-I) TF (a member of the Apicomplexan Apetala 2 (ApiAP2) TF). *Plasmodium falciparum* Apetala 2 – Invasion (*PfAP2-I*) TF is a 183 KDa protein located on chromosome 10 of the *Plasmodium* genome whose primary function is to regulate red blood cell (RBC) invasion genes [15]. *PfAP2-I* also targets promoters of nucleosome- and chromatin-related genes, cell-cycle-related genes, and genes associated with vesicle transport and host-cell remodeling [14, 16]. It contains three AP2 domains, and only the third AP2 domain is essential for regulating a subset of genes involved in red blood cell (RBC) invasion [15, 16]. *PfAP2-I* associates with several chromatin associated proteins, including the *Plasmodium* bromodomain protein 1 (*PfBDP1*) and that complex formation is associated with transcriptional regulation. *PfAP2-I* represents a potential new antimalarial therapeutic target as a critical regulator of red blood cell invasion genes [17]. It is important to study and understand its functioning and determine drugs that can inhibit its activity. This is very important for disrupting the parasite cycle in the human host, and to design effective therapies that can augment efficacies of existing antimalarials. This study aims to identify small molecules and validate the inhibitory potential of these small molecules against AP2-I regulatory action in *Plasmodium falciparum*.

## Methods

### *PfAP2-I* structure prediction

The experimental 3D structure of *PfAP2-I* was modeled because it is not available in the Protein Data Bank (PDB) [18] as well as UniProt Knowledge base (UniProtKB) [19]. The protein ID of the target (*Plasmodium falciparum* Apetala 2-Invasion 3D7 strain) was retrieved from National Centre for Biotechnology Information (NCBI) with the accession number PF3D7\_1007700.

Afterward, the protein ID was submitted to the SWISS-MODEL web server [20] to develop a homology model with sufficient query sequence coverage and sequence identity. The confident match to a protein of known structure was below 40%, so comparative modeling of *PfAP2-I* could not be done. The 3D structure of *PfAP2-I* was then modeled on both the I-TASSER server (<http://zhanglab.dcmf.med.umich.edu/I-TASSER>) [21] and ROBETTA Baker server (<http://rosetta.bakerlab.org>) [22] using RoseTTAFold.

RoseTTAFold is the default option that uses a deep learning-based modeling method. This method outperforms every other way for protein structure modeling on the ROBETTA Baker server. The most reliable 3D structure was selected based on the confidence value. The confidence values are usually between 0.00 (bad) and 1.00 (good), and the higher the number, the higher the reliability of the predicted structure.

### Structure validation of modeled protein

PROCHECK [23] and ERRAT [24] on UCLA-DOE LAB – SAVES v6.0 were used to check for the quality of the modeled 3D structure of *PfAP2-I* generated on the ROBETTA Baker Lab. The .pdb file format of the modeled *PfAP2-I* was uploaded on the UCLA-DOE LAB – SAVES v6.0 site for this structure validation. The .pdb file format of the modeled *PfAP2-I* was uploaded on the server to obtain the overall quality factor from ERRAT and Ramachandran plot and the Ramachandran plot statistics from PROCHECK. The overall quality factor is expressed as the percentage of protein for which the calculated error value falls below the 95% rejection limit. Good high-resolution structures usually produce values around 95% or higher.

The Ramachandran plot is used in accessing the quality of a modeled protein or an experimental structure, while the Ramachandran plot statistics provide information on the total number of amino acid residues found in the favorable, allowed, and disallowed regions [23].

### Active site prediction of AP2-I modeled structure and *PfBDP1*

The crystal structure of *PfBDP1* was retrieved from Protein Data Bank (PDB) ([www.rcsb.org/structure/7M97](http://www.rcsb.org/structure/7M97)) [18]. The active sites of modeled *PfAP2-I* 3D7 and *PfBDP1* structure were predicted using Computed Atlas of Surface Topography of proteins (CASTp) 3.0. [25] and *ConCavity* [26]. The Computed Atlas of Protein Surface Topography (CASTp) is an online service for identifying, defining, and quantifying certain geometric and topological features of protein structures such as surface pockets, interior cavities, and cross channels (Dundas et al., 2006), while *ConCavity* is an online service used for predicting Protein-Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure and works based on confidentiality score (C-score). C-score is a confidence score of the predicted binding site. C-score values range between 0-1, where a higher score indicates a more reliable prediction. The modeled *PfAP2-I* and *PfBDP1* 3D7 structure were submitted on the server. The necessary amino acids for binding interactions predicted by the two servers were compared to determine the similarity between the two predicted active sites.

### Pharmacophore modeling

A pharmacophore model using the prepared modeled 3D structure of *PfAP2-I* TF was designed using pharmit server [28]. Pharmit server is a collection of built-in databases such as Molprot, ChEMBL, ZINC, and PubChem. It contains millions of chemical compounds that can be used to screen drug-like compounds against a given protein [29]. The Pharmit server is based on a pharmacophore model using the AutoDock Vina scoring function [30]. A control ligand (3W7 from COACH server) was selected for the screening [31], and both the modeled protein and control ligand were loaded into the Pharmit Server. The pharmacophore model was built using six features, i.e., one hydrogen donor, two hydrogen acceptors, one aromatic, and two hydrophobic. The pharmit filters hit screening for the pharmacophore modeling were set using the Lipinski rule of 5 to minimize the results significantly and obtain the best possible inhibitors out of millions of drug-like compounds.

Pharmit filter hit screening based on Lipinski rule of 5

$0 \leq \text{Molecular weight} \leq 500$

$0 \leq \text{Rotatable bonds} \leq 10$

$0 \leq \text{LogP} \leq 5$

$0 \leq \text{Polar Surface Area} \leq 140$

$0 \leq \text{Molecular weight} \leq 500$

$0 \leq \text{Hydrogen Bond Acceptor} \leq 10$

$0 \leq \text{Hydrogen Bond Donor} \leq 5$

### ***Protein and Ligand Preparation***

The modeled protein structure was defined as a receptor while the complexed ligands were removed using Chimera software [32]. Furthermore, the protein was prepared by computing Gasteiger charges, adding polar hydrogens, and merging the nonpolar hydrogens using AutoDockTools4.2.6. [33]. In addition, OpenBabel software [34] was used to convert the .pdb files to the AutoDock docking format (.pdbqt), which was further used for the docking simulation.

### ***Virtual screening analysis***

The virtual screening of compounds was carried out using AutoDock Vina, an accessible graphical user interface (GUI) for the AutoDock 4.2 program [35]. The grid box was constructed using 80, 80, and 91 pointing in x, y, and z directions, respectively, with a grid point spacing of 0.375 Å. The center grid box is of 108.636 Å, 73.665 Å, and 158.751 Å around THR508A, TRP510A, LYS512A, THR514A, THR515A, GLU516A, GLU520A, TYR521A, LEU522A, GLN535A, VAL554A, LYS555A, TYR557A, GLY558A, GLN561A, ALA562A, HIS585A, VAL586A, HIS587A, GLY588A, ARG590A, LYS591A, VAL593A, ASP594A, THR598A. These amino acids were selected based on the CASTp and *Concavity* result. The top ten (10) hits against PfAP2-I were then generated and were ranked according to their binding affinities to verify the ligand-binding sites. The top six (6) best-docked compounds from PfAP2-I docking analysis were also docked against PfBDP1 active site. Post-screening analyses were conducted using AutoDockTools, and LigPlot [36].

### ***In silico drug-likeness and toxicity predictions***

The *in silico* drug-likeness and toxicity predictions of the designed ligands were carried out using the Swiss ADME predictor [37] and OSIRIS Property Explorer [38]. SwissADME predictor provides information on the Oral bioavailability, Physicochemical properties, Lipophilicity, Water solubility, Pharmacokinetics, Druglikeness, and Medicinal chemistry of the compounds [39]. OSIRIS Property Explorer program, on the other hand, provides information on a compound's toxicity and determines parameters such as Molecular weight, Consensus lipophilicity (cLogP), Total polar surface area (TPSA), Solubility, Drug-likeness, and Drug score, as well as the mutagenic, tumorigenic, irritant and reproductive risks [40].

Drug-likeness is a criterion for determining if a pharmacological substance possesses the characteristics of an orally active drug [41]. The Lipinski rule of five is an established concept upon which drug-likeness is based. The law states that for a compound to exhibit drug-likeness and to avoid poor absorption or permeation, the combination must not possess more than 5H-bond donors, more than 10H-bond acceptors, molecular weight must not be greater than 500, and the calculated LogP (cLogP) must not be greater than 5 [41].

Another parameter used to select compounds as drug candidates are drug score. A high drug score value signifies a high probability of the compound being considered a drug candidate [43].

## **Results**

### ***PfAP2-I structure prediction***

PfAP2-I structure prediction result from the I-TASSER server shows four predicted models. The models were predicted based on C-score, Exp. TM-Score, Exp.RMSD, No.of decoys and Cluster density, and the result for each model is shown in Table 1. PfAP2-I structure prediction result from the ROBETTA Baker server showed five predicted models using RoseTTAFold and model one (1) was prioritized based on its highest C-score. RoseTTAFold is the default option and the most accurate method on the ROBETTA Baker server that uses a deep learning-based modeling method. For the RoseTTAFold, the confidence corresponds to the predicted local distance difference test (LDDT) using DeepAccNet.

### ***Structure validation of modeled protein***

The PfAP2-I 3D modeled structure results from both the I-TASSER server and ROBETTA Baker server were validated using ERRAT and PROCHECK on UCLA-DOE LAB – SAVES v6.0. Model 1 of the PfAP2-I 3D structure from the ROBETTA Baker server showed the best result after validation. Errat value for model 1 of the PfAP2-I 3D structure from the ROBETTA Baker server was 96.827, and good resolution values usually produce around 95% or higher (Fig. 3). The Ramachandran plot statistics implied that the model 1 3D structure of PfAP2-I from ROBETTA Baker server has 90.2% of its residues in the most favored regions, 8.5% of its residues in additional allowed regions, 0.4% of its residues in the generously allowed regions, and 0.9% of its residues in disallowed regions of the Ramachandran plot. A good quality model is expected to have 90% of its residues in the most favored region. This also validates that the 3D modeled structure is a good quality model (Fig. 3)

### Active site prediction of AP2-I modeled structure

Using CASTp 3.0 active site prediction tool, a total of 102 pockets were generated, and pocket ID 2 was selected as the preferred active site with an area (SA) of 707.191 and a volume (SA) of 1193.757 (Fig. 4a). For *ConCavity*, rank one was selected with the highest C-score of 0.66. Pocket 2 from CASTp 3.0 and rank one of *ConCavity* were chosen as the more favorable sites for the docking analyses due to the similarities observed from the comparison of results. Pocket 2 from CASTp 3.0 has 25 amino acid residues in the active sites for the modeled protein - THR508A, TRP510A, LYS512A, THR514A, THR515A, GLU516A, GLU520A, TYR521A, LEU522A, GLN535A, VAL554A, LYS555A, TYR557A, GLY558A, GLN561A, ALA562A, HIS585A, VAL586A, HIS587A, GLY588A, ARG590A, LYS591A, VAL593A, ASP594A, THR598A.

### Active site prediction of PfBDP1 3D structure

Using CASTp 3.0 active site prediction tool, a total of 11 pockets were generated, and pocket ID 1 was selected as the preferred active site with an area (SA) of 147.912 and a volume (SA) of 74.504 (Fig. 4b). The amino acid residues present in chain A of pocket 1 are: ILE355A, PHE356A, LYS358A, LEU359A, VAL360A, GLN365A, CYS367A, TYR370A, PRO377A, MET378A, 379SERA, ILE405A, ASN408A, CYS409A, TYR412A, ASN413A, and VAL419A.

### Pharmacophore Based Virtual Screening result

Hit screening of drug-like like compounds from nine (9) databases (ChEMBL, ChemDiv, ChemSpace, MCULE, MCULE-ULTIMATE, MolPort, NCI open repository, Lab Network, Zinc) on the Pharmit Server Engine gave a total of eight thousand, one hundred and fifteen (8,656) drug-like compounds (Table 2). These eight thousand, six hundred and fifty-six (8,656) drug-like compounds were used for the docking analysis as well as the reference compound CHEMBL3359262) and the standard drug (chloroquine) against *PfAP2-I*, and the top ten (10) hits were identified based on the highest-ranked auto-dock score from the docking analysis (Table 3). The result of the top six (6) best-docked compounds docked against *PfBDP1*'s active site is shown in Table 4.

### Post-docking analysis

Six (6) out of the ten (10) compounds were selected for post-docking (LigPlot webserver) analysis because three compounds were repetitions (compounds in \*), and no information was found for one (the compound in #) from PubChem. The 2D and 3D model interaction between best hit compound (ZINC97139187) and *PfAP2-I* is shown in Fig. 6 while that of *PfBDP1* is shown in Fig.7. The hydrogen bonds interactions as well as the bond lengths of two best hit compounds from the docking analysis against *PfAP2-I* and *PfBDP1* are shown in Table 5.

### In silico results of risks and drug-likeness of six compounds

ADMET screening and toxicity testing results of the six (6) drug-like compounds are shown in Table 6 and 7 while the oral bioavailability radar results for the six (6) drug-like compounds including that of the reference compound (CHEMBL3359262) and standard drug (chloroquine) is shown in Fig. 8.

## Discussion

Previous genome-wide analysis by ChIP-seq show that *PfAP2-I* interacts with a specific DNA motif in the promoters of target genes [17]. Although *PfAP2-I* contains three AP2 DNA-binding domains, only one is required for binding of the target genes during blood stage development [17]. *PfAP2-I* also associates with several chromatin associated proteins, including the *Plasmodium falciparum* bromodomain protein (*PfBDP1*) and that complex formation is associated with transcriptional regulation [17, 44]. A study on *PfBDP1* regulation of invasion gene expression shows that *PfBDP1* binds to chromatin at transcriptional start sites of invasion-related genes and directly controls their expression [44]. Conditional *PfBDP1* knock-down causes a dramatic defect in parasite invasion and growth and results in transcriptional down-regulation of multiple invasion-related genes at a time point critical for invasion. Also, *PfBDP1* overexpression enhances expression of these same invasion related genes. *PfBDP1* may also be independently recruited to chromatin through other DNA-binding proteins. These findings collectively show that *PfBDP1* critically coordinates expression of invasion genes and indicates that targeting *PfBDP1* could be an invaluable tool in malaria eradication [44].

In this current study, molecular docking analysis of 8,656 compounds was carried out to determine the inhibitory potential of small molecules against *PfAP2-I* TF and a bromodomain protein *PfBDP1*. These compounds were used to prepare the ligand library and screened against the prepared 3D model structure of *PfAP2-I* and *PfBDP1*. The top ten (10) drug-like compounds were identified based on the docking analysis's highest-ranked auto-dock score. Six (6) out of the ten (10) compounds were selected for post docking analysis because three compounds were repetitions (compounds in \*), and no information was found for one (the compound in #) from PubChem. The six compounds exhibited dock scores between -9.9 and -10.2 kcal/mol, and had lower binding energies when compared to the reference compound (CHEMBL3359262) and the standard drug (chloroquine) with binding energies of -8.4 and -4.2 kcal/mol respectively for *PfAP2-I*. *PfBDP1*, as a co-regulator of *PfAP2-I* was also subjected to docking analysis in this study to determine the inhibitory potential of the six best hits from *PfAP2-I* docking. The six selected compounds exhibited dock scores between -8.5 and -9.4 kcal/mol, and had lower binding energies when compared to the reference compound (CHEMBL3359262) and the standard drug (chloroquine) with binding energies of -8.3 and -6.1 kcal/mol respectively when docked against *PfBDP1*.

ADMET screening of the six (6) drug-like compounds were based on parameters such as the bioavailability radar, lipophilicity, water-solubility, pharmacokinetics, drug-likeness, and medicinal chemistry. The oral bioavailability radar summarily describes the degree of drug-likeness of a molecule first using six properties (lipophilicity, size, polarity, insolubility, saturation, and flexibility) (Fig. 8). For each of the properties, the pink area represents the physicochemical space with an optimal range of lipophilicity (XLOGP3) between -0.7 and +5.0, size (molecular weight) between 150 and 500g/mol, polarity (total polar surface area) between 20 and 130 Å<sup>2</sup>, solubility (log S) not more than 6, saturation not less than 0.25 and flexibility, not more than nine rotatable

bonds. ZINC97139187, 57410073, 154861216, and 57405339 fell within the physicochemical space for all six properties and can be said to be orally bioavailable, ChEMBL3923620, and MCULE-6567089130 had a deviation in the degree of saturation outside the physicochemical space.

Table 7 and 8 showed that all the compounds were within the minimum and maximum acceptable range. The six (6) compounds had molecular weights less than 500. This implies that when these compounds are administered as drugs, they are likely to be absorbed and reach the site of action [45]. The numbers of rotatable bond (NRB), hydrogen bond acceptors (NHA), and hydrogen bond donors (NHD) in the six compounds (Table 7) are in accordance with Lipinski's rule of five [42].

Lipophilicity is an essential property in drug discovery. Multiple predictors (iLOGP, XLOGP3, WLOGP, MLOGP and SILICOS-IT) were used to generate a consensus estimation of lipophilicity to increase the prediction accuracy. Consensus LogP values of less than five were found in all substances, including the standard medication (chloroquine), indicating good absorption and penetration across cell membranes [45].

Solubility of a molecule greatly facilitates major drug development activities, especially the ease of handling and drug formulation. It is a significant property influencing absorption and delivery of a sufficient quantity of active ingredients in the small volume of pharmaceutical dosage [46]. Five compounds (ZINC97139187, 57410073, 154861216, MCULE-6567089130, and 57405339) were moderately soluble in water with LogS (ESOL) values ranging from -5.42 to -4.60, while one (ChEMBL3923620) of the compounds was soluble in water with a -3.41 LogS (ESOL) value.

Pharmacokinetic properties such as gastrointestinal (GI) absorption, Blood-Brain Barrier (BBB), and CYP 450 enzymes can be used to evaluate individual ADME behavior of small molecules. It has been suggested that CYP can process small molecules synergistically to improve the protection of tissues and organisms [47]. It is estimated that 50 to 90% of therapeutic molecules are substrates of five major isoforms (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4) [48]. All six compounds have a high gastrointestinal (GI) absorption and cannot cross the Blood-Brain Barrier (BBB), except 154861216. Four of the compounds are suitable inhibitors of CYP450 enzymes, while two are not (Table 5).

The selected compounds' physicochemical properties and toxicity risks were carried out using *Osiris property explorer*. Molecular weight, solubility prediction, total polar surface area (TPSA) ( $\text{\AA}^2$ ), drug-likeness, drug score, mutagenic tumorigenic, irritant, and reproductive properties were examined predicted. Compounds with high molecular weights above 500 g/mol are less likely to be absorbed and, therefore, less likely to reach the destination of action. The molecular weights of the six compounds were found to be less than 500. Drug-likeness may be defined as a complex balance of molecular properties and structural features that determine whether a particular molecule is similar to the known drugs. A positive value for drug-likeness means that the molecule contains predominantly fragments frequently present in commercial drugs [49]. Four (4) compounds, ZINC97139187, 57410073, 154861216, and MCULE-6567089130, had positive value for the drug-likeness property.

The drug score (ds) is a contribution calculated directly from parameters of the Partition coefficient (cLogP), solubility(clogS), molecular weight (Mol. Wt), drug-likeness, and toxicity risk within one good practical value [50]. The higher the drug score, the better the chance to be a drug candidate. The drug score values such as 1.0, 0.8, and 0.6 are associated with no risk, medium risk, and high risk, respectively. Among the six compounds, ZINC97139187 had the highest drug score value of 0.63 (more elevated than that of chloroquine of 0.25), fell within the medium-risk range, and may be used as a drug molecule. In general, the drug score values of the five compounds (ZINC97139187, 57410073, 154861216, MCULE-6567089130, and 57405339) with range 0.26-0.63 were more significant than that of Chloroquine. Only one compound (ChEMBL3923620) was below that of Chloroquine with a value of 0.14. This compound (ChEMBL3923620) also possessed medium risk mutagenic, low risk tumorigenic, high-risk irritant and low reproductive toxicity risks. Another compound (57410073) also showed low risk mutagenic, high risk tumorigenic, low-risk irritant, and low reproductive toxicity risks. All other compounds (ZINC97139187, 154861216, MCULE-6567089130, and 57405339) had low risk mutagenic, low risk tumorigenic, low-risk irritant, and low-risk reproductive toxicity risks.

## Conclusion

The good, estimated binding energies and drug score values observed for compound ZINC97139187 against *Pf*AP2-I and compound MCULE-6567089130 against *Pf*BDP1 suggest that they can be considered possible *Pf*AP2-I and *Pf*BDP1 inhibitors. Further pre-clinical experimental validations should be carried out to ascertain the efficacy of these predicted best hits.

## List Of Abbreviations

TF: Transcription Factor

IDC: Intra-erythrocytic Development Cycle

RBC: Red Blood Cell

*Pf*AP2-I: *Plasmodium falciparum* Apetala 2 – Invasion

*Pf*BDP1: *Plasmodium* bromodomain protein 1

PDB: Protein Data Bank

## Declarations

Ethics approval and consent to participate

Not applicable.

#### Consent to publish

All authors read and agreed to publish this study.

#### Availability of data and materials

The supplementary data relevant to this study is provided

#### Competing interests

The authors declare that they have no competing interests.

#### Funding

This study received research funding from Covenant Applied Informatics and Communication Africa Centre of Excellence (CAPIC-ACE) under the World Bank Africa Centre of Excellence (ACE Impact) project and publication funding support from Covenant University Centre for Research, Innovation and Discovery (CUCRID), Nigeria.

#### Authors' Contributions

All authors contributed significantly to the study: EFA and DO designed and led the study, GO, TD, II, EO, and EEJI and participated in the conduct and analysis, EFA, EEJI, GO and TD participated in the conduct, review, and correction of the final manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

Not applicable.

## References

1. Gregson A, Plowe C v. Mechanisms of Resistance of Malaria Parasites to Antifolates. *Pharmacological Reviews*. 2005;57:117–45.
2. Nsanjabana C. Tropical Medicine and Infectious Disease Resistance to Artemisinin Combination Therapies (ACTs): Do Not Forget the Partner Drug! 2019. <https://doi.org/10.3390/tropicalmed4010026>.
3. Doshi K, Pandya N, Datt M. In silico assessment of natural products and approved drugs as potential inhibitory scaffolds targeting aminoacyl-tRNA synthetases from Plasmodium. *3 Biotech*. 2020;10.
4. Joseph Sahayarayan J, Soundar Rajan K, Nachiappan M, Prabhu D, Guru Raj Rao R, Jeyakanthan J, et al. Identification of potential drug target in malarial disease using molecular docking analysis. *Saudi journal of biological sciences*. 2020;27:3327–33.
5. Sheikh IA, Jiffri EH, Ashraf GM, Kamal MA, Beg MA. Structural studies on inhibitory mechanisms of antibiotic, corticosteroid and catecholamine molecules on lactoperoxidase. *Life sciences*. 2018;207:412–9.
6. Kassiou M, Brogi S, Ramalho TC, Kuca K, Medina-Franco JL, Valko M, et al. Editorial: In silico Methods for Drug Design and Discovery. *Frontiers in Chemistry* | [www.frontiersin.org](http://www.frontiersin.org). 2020;1:612.
7. Vardhan S, Sahoo SK. In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19. *Computers in Biology and Medicine*. 2020;124.
8. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT. Computational approaches in target identification and drug discovery. *Computational and structural biotechnology journal*. 2016;14:177–84.
9. Siwo GH, Smith RS, Tan A, Button-Simons KA, Checkley LA, Ferdig MT. An integrative analysis of small molecule transcriptional responses in the human malaria parasite Plasmodium falciparum. *BMC genomics*. 2015;16.
10. Papavassiliou KA, Papavassiliou AG. Transcription Factor Drug Targets. *Journal of cellular biochemistry*. 2016;117:2693–6.
11. Lambert M, Jambon S, Depauw S, David-Cordonnier MH. Targeting Transcription Factors for Cancer Treatment. *Molecules (Basel, Switzerland)*. 2018;23.
12. Kohzaki H, Murakami Y. Transcription factors and DNA replication origin selection. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 2005;27:1107–16.
13. Toenhake CG, Bártfai R. What functional genomics has taught us about transcriptional regulation in malaria parasites. *Briefings in functional genomics*. 2019;18:290–301.
14. Toenhake CG. Transcription factors and gene regulatory elements in the human malaria parasite. *Plasmodium falciparum*. undefined. 2020.
15. Campbell TL, de Silva EK, Olszewski KL, Elemento O, Llinás M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS pathogens*. 2010;6.
16. Toenhake CG, Fraschka SAK, Vijayabaskar MS, Westhead DR, van Heeringen SJ, Bártfai R. Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying Plasmodium falciparum Blood-Stage Development. *Cell host & microbe*. 2018;23:557-569.e9.
17. Santos JM, Josling G, Ross P, Joshi P, Orchard L, Campbell T, et al. Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor. *Cell host & microbe*. 2017;21:731-741.e10.

18. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, di Costanzo L, et al. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*. 2019;47:D464–74.
19. Bateman A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. 2019;47 Database issue:D506.
20. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*. 2018;46:W296–303.
21. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods*. 2015;12:7–8.
22. Chivian D, Kim DE, Malmström L, Bradley P, Robertson T, Murphy P, et al. Automated prediction of CASP-5 structures using the Robetta server. *Proteins*. 2003;53 Suppl 6 SUPPL. 6:524–33.
23. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein-structure coordinates . 2012;;684–7.
24. Dym O, Eisenberg D, Yeates TO. Detection of errors in protein models. *International Tables for Crystallography*. 2006;;520–30.
25. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*. 2018;46:W363–7.
26. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure. *PLoS Comput Biol*. 2009;5:1000585.
27. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. <https://doi.org/10.1093/nar/gkl282>.
28. Sherma A, Yogavel M, Sharma A. Structural and functional attributes of malaria parasite diadenosine tetraphosphate hydrolase. *Scientific reports*. 2016;6.
29. DS W, YD F, AC G, EJ L, A M, JR G, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*. 2018;46.
30. Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. *Nucleic acids research*. 2016;44:W442–8.
31. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics (Oxford, England)*. 2013;29:2588–95.
32. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*. 2004;25:1605–12.
33. Singh IV, Mishra S. Molecular Docking Analysis of Pymethamine Derivatives with Plasmodium falciparum Dihydrofolate Reductase. *Bioinformation*. 2018;14:232–5.
34. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *Journal of cheminformatics*. 2011;3.
35. Morris GM, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*. 2009;30:2785–91.
36. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of chemical information and modeling*. 2011;51:2778–86.
37. Lohidashan K, Rajan M, Ganesh A, Paul M, Jerin J. Pass and Swiss ADME collaborated in silico docking approach to the synthesis of certain pyrazoline spacer compounds for dihydrofolate reductase inhibition and antimalarial activity. *III Bangladesh Journal of Pharmacology III*. 2018;13:23–9.
38. Torres E, Moreno E, Ancizu S, Barea C, Galiano S, Aldana I, et al. New 1,4-di-N-oxide-quinoxaline-2-ylmethylene isonicotinic acid hydrazide derivatives as anti-Mycobacterium tuberculosis agents. *Bioorganic & medicinal chemistry letters*. 2011;21:3699–703.
39. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*. 2017;7.
40. Mahato S, Singh A, Rangan L, Jana CK. Synthesis, In silico studies and In vitro evaluation for antioxidant and antibacterial properties of diarylmethylamines: A novel class of structurally simple and highly potent pharmacophore. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences*. 2016;88:202–9.
41. Egbert M, Whitty A, Keserü GM, Vajda S. Why Some Targets Benefit from beyond Rule of Five Drugs. *Journal of medicinal chemistry*. 2019;62:10005–25.
42. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*. 2001;46:3–26.
43. Behrouz S, Soltani Rad MN, Taghavi Shahraki B, Fathalipour M, Behrouz M, Mirkhani H. Design, synthesis, and in silico studies of novel eugenylpropanol azole derivatives having potent antinociceptive activity and evaluation of their  $\beta$ -adrenoceptor blocking property. *Molecular diversity*. 2019;23:147–64.
44. Josling GA, Petter M, Oehring SC, Gupta AP, Dietz O, Wilson DW, et al. A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. *Cell host & microbe*. 2015;17:741–51.
45. Wu CY, Benet LZ. Predicting drug disposition via application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharmaceutical research*. 2005;22:11–23.
46. Savjani KT, Gajjar AK, Savjani JK. Drug solubility: importance and enhancement techniques. *ISRN pharmaceuticals*. 2012;2012:1–10.
47. Di L. The role of drug metabolizing enzymes in clearance. *Expert opinion on drug metabolism & toxicology*. 2014;10:379–93.
48. Wolf CR, Smith G, Smith RL. Science, medicine, and the future: Pharmacogenetics. *BMJ (Clinical research ed)*. 2000;320:987–90.
49. Ursu O, Rayan A, Goldblum A, Oprea TI. Understanding drug-likeness. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2011;1:760–81.
50. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*. 2006;26:531–68.



Table 1. I-TASSER *PfAP2-I* structure prediction result

Name	C-score	Exp.TM-Score	Exp.RMSD	No.of decoys	Cluster density
Model1	0.03	0.72+0.11	9.7+4.6	600	0.5924
Model2	-2.64			251	0.0410
Model3	-3.40			84	0.0191
Model4	-3.93			60	0.0113
Model5	-3.87			60	0.0120

Table 2: Pharmacophore based virtual screening of compounds from nine (9) databases on the Pharmit server

Pharmit database	Conformers	Molecules	Hits
a. ChEMBL25	23,136,925	1,752,844	182
b. ChemDiv	21,562,497	1,456,120	77
c. ChemSpace	250,205,463	50,181,678	1,719
d. MCULE	223,460,579	45,257,086	1,716
e. MCULE-ULTIMATE	378,880,344	126,471,502	2,018
f. MolPort	114,798,054	8,015,098	1,534
g. NCI Open Chemical Repository	574,117	52,237	-
h. LabNetwork	22,051,020	1,794,286	138
i. Zinc	123,399,574	13,190,317	1,272
<b>Total</b>			<b>8,656</b>

Table 3: The top 10 best-docked compounds from the Pharmit server ranked based on the Auto-dock score against *PfAP2-I* active site

Docked ligand	Pharmit IDs	Pubchem IDs	Molecular formula	Binding affinity Kcal/mol
a.	ZINC97139187*	87052587*	C <sub>28</sub> H <sub>24</sub> N <sub>6</sub> O <sub>3</sub> *	-10.2
b.	CHEMBL3923620	123492565	C <sub>19</sub> H <sub>14</sub> FN <sub>9</sub> O <sub>2</sub>	-10.1
c.	57410073	57410073	C <sub>29</sub> H <sub>29</sub> N <sub>5</sub> O <sub>2</sub>	-10.0
d.	ChemDiv-S022-1700*	87052587*	C <sub>28</sub> H <sub>24</sub> N <sub>6</sub> O <sub>3</sub> *	-10.0
e.	154861216	154861216	C <sub>26</sub> H <sub>30</sub> N <sub>4</sub> O <sub>2</sub>	-10.0
f.	MCULE-7146940834*	87052587*	C <sub>28</sub> H <sub>24</sub> N <sub>6</sub> O <sub>3</sub> *	-10.0
g.	MCULE-6816682540#	-	-	-10.0
h.	MolPort-035-696-264*	87052587*	C <sub>28</sub> H <sub>24</sub> N <sub>6</sub> O <sub>3</sub> *	-9.9
i.	MCULE-6567089130	135500213	C <sub>20</sub> H <sub>13</sub> N <sub>7</sub> O	-9.9
j.	57405339	57405339	C <sub>28</sub> H <sub>31</sub> N <sub>7</sub> O	-9.9
k.	Reference compound (CHEMBL3359262)	52934178	C <sub>22</sub> H <sub>25</sub> N <sub>5</sub> O <sub>2</sub>	-8.4
l.	Chloroquine	2719	C <sub>18</sub> H <sub>26</sub> ClN <sub>3</sub>	-4.2

\*- repeated compounds

# - was not found on pubchem

Table 4: The top six (6) best-docked compounds from the Pharmit server docked against *PfBDP1* active site

Docked ligand	Pharmit IDs	Pubchem IDs	Molecular formula	Binding affinity Kcal/mol
a.	MCULE-6567089130	135500213	C <sub>20</sub> H <sub>13</sub> N <sub>7</sub> O	-9.4
b.	57405339	57405339	C <sub>28</sub> H <sub>31</sub> N <sub>7</sub> O	-9.3
c.	57410073	57410073	C <sub>29</sub> H <sub>29</sub> N <sub>5</sub> O <sub>2</sub>	-9.1
d.	ZINC97139187	87052587	C <sub>28</sub> H <sub>24</sub> N <sub>6</sub> O <sub>3</sub>	-9.0
e.	CHEMBL3923620	123492565	C <sub>19</sub> H <sub>14</sub> FN <sub>9</sub> O <sub>2</sub>	-8.8
f.	154861216	154861216	C <sub>26</sub> H <sub>30</sub> N <sub>4</sub> O <sub>2</sub>	-8.5
g.	Reference compound (CHEMBL3359262)	52934178	C <sub>22</sub> H <sub>25</sub> N <sub>5</sub> O <sub>2</sub>	-8.3
h.	Chloroquine	2719	C <sub>18</sub> H <sub>26</sub> ClN <sub>3</sub>	-6.1

Table 5. Hydrogen bonds interaction between the best hit compounds against *Pf*AP2-I and *Pf*BDP1 based on binding affinity using LigPlot

	Pharmit IDs	Pubchem IDs	Protein	Hydrogen bonds and the bond lengths
a.	ZINC97139187	87052587	<i>Pf</i> AP2-I	Lys591A (3.17 Å)
b.	MCULE-6567089130	87052587	<i>Pf</i> BDP1	Asn413A (2.81 Å), Ile355A (2.79 Å)

Table 6. SwissADME prediction of ADME properties of six compounds in comparison with chloroquine

Compounds	Physicochemical properties			Lipophilicity cLogP	Water solubility LOG (ESOL)	Pharmacokinetics					
	NRB	NHA	NHD			GI Absorption	BBB Permeant	CYP1A2 Inhibitor	CYP2C19 Inhibitor	CYP2C9 Inhibitor	CYP2D6 Inhibitor
a. ZINC97139187	4	5	2	1.44	-4.95	High	No	Yes	Yes	Yes	Yes
b. CHEMBL3923620	4	9	2	0.77	-3.41	High	No	No	No	No	No
c. 57410073	6	4	2	3.84	-5.42	High	No	No	Yes	Yes	Yes
d. 154861216	5	3	1	4.14	-4.78	High	Yes	No	Yes	Yes	Yes
e. MCULE-6567089130	1	4	3	2.11	-4.60	High	No	Yes	No	No	No
f. 57405339	6	5	1	2.76	-5.30	High	No	No	Yes	Yes	Yes
g. Reference compound (CHEMBL3359262)	3	4	1	2.57	-3.94	High	No	No	No	Yes	Yes
h. Chloroquine	1	2	8	4.01	-4.55	High	Yes	Yes	No	No	Yes

Table 7. OSIRIS Property Explorer prediction of physicochemical properties and toxicity risks of six compounds in comparison with chloroquine

Compounds	Physicochemical properties					Toxicity risks			
	Molecular weight (g/mol)	Solubility prediction	TPSA (Å <sup>2</sup> )	Drug likeness	Drug score	Mutagenic	Tumorigenic	Irritant	Reproductive
a. ZINC97139187	492.5	-4.3	108.38	6.76	0.63	Low Risk	Low Risk	Low Risk	Low Risk
b. ChEMBL3923620	419.4	-6.48	139.76	-1.67	0.14	Medium risk	Low Risk	High Risk	Low Risk
c. 57410073	479.6	-6.08	90.98	3.86	0.26	Low Risk	High Risk	Low Risk	Low Risk
d. 154861216	430.5	-4.16	69.30	2.04	0.58	Low Risk	Low Risk	Low Risk	Low Risk
e. MCULE-6567089130	367.4	-7.86	117.75	4.37	0.47	Low Risk	Low Risk	Low Risk	Low Risk
f. 57405339	481.6	-3.71	82.94	-0.27	0.47	Low Risk	Low Risk	Low Risk	Low Risk
g. Reference compound (ChEMBL3359262)	391.0	-2.97	83.88	8.43	0.82	Low Risk	Low Risk	Low Risk	Low Risk
h. Chloroquine	319.9	-4.06	28.16	7.39	0.25	High Risk	Low Risk	High Risk	Low Risk

## Figures

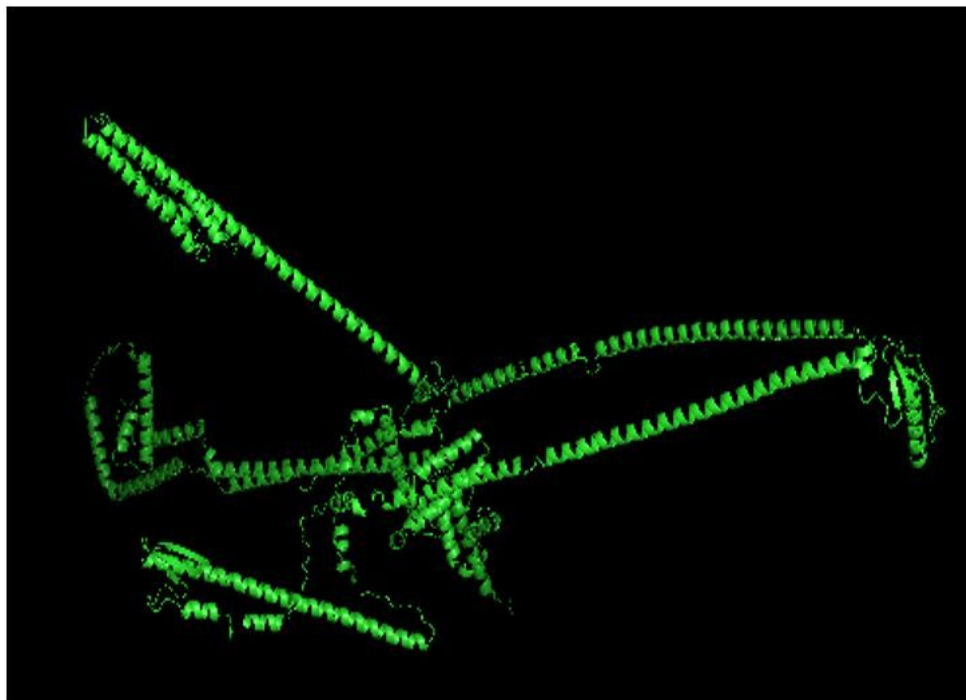
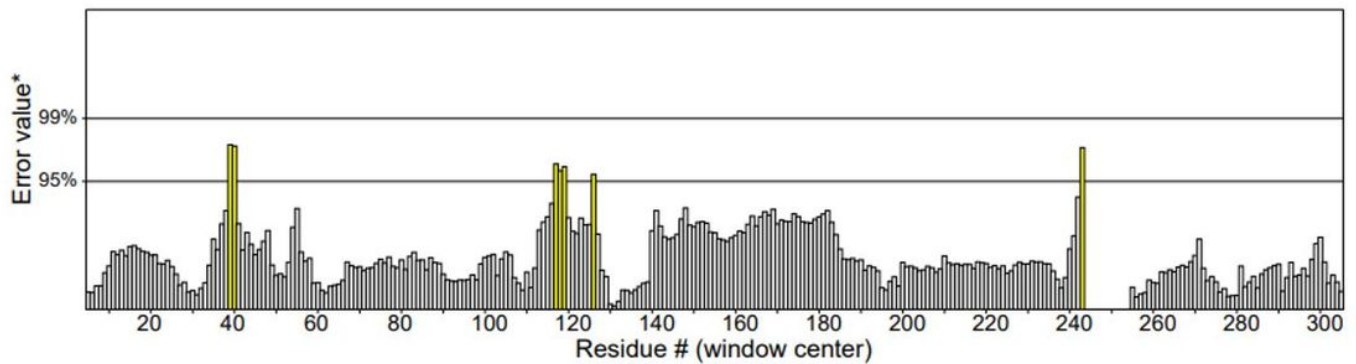


Figure 1

3D modeled structure of *PfAP2-I* TF from ROBETTA Baker server

Program: ERRAT2  
File: model01.pdb  
Chain#:A  
Overall quality factor\*\*: 96.827

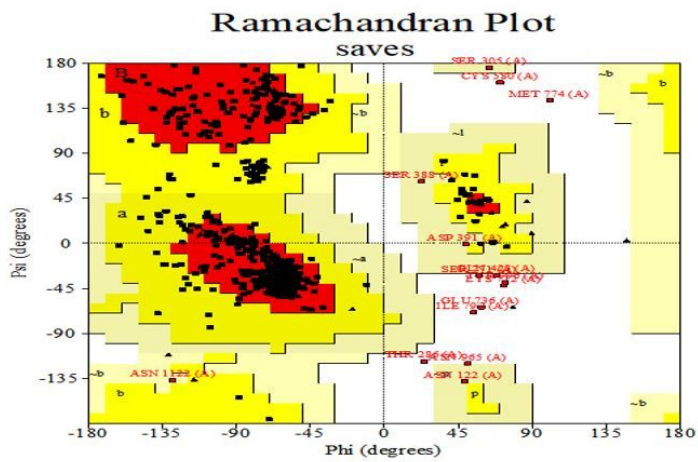


\*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

\*\*Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3Å) the average overall quality factor is around 91%.

Figure 2

ERRAT structure validation value of *PfAP2-I* modeled structure



#### Plot statistics

Residues in most favoured regions [A,B,L]	1017	90.2%
Residues in additional allowed regions [a,b,l,p]	96	8.5%
Residues in generously allowed regions [~a,~b,~l,~p]	5	0.4%
Residues in disallowed regions	10	0.9%
-----		
Number of non-glycine and non-proline residues	1128	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	48	
Number of proline residues	22	
-----		
Total number of residues	1200	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 3

Ramachandran plot statistic validation of *PfAP2-I* modeled structure

a.



b.

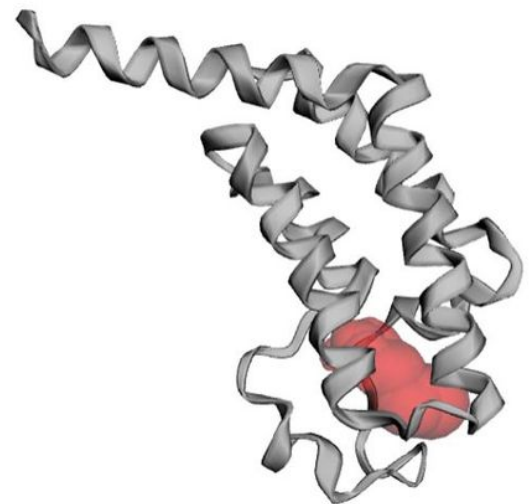


Figure 4

The binding pocket of 3D structure *Pf*AP2-I (a) and *Pf*BDP1 (b) as computed using CASTp 3.0

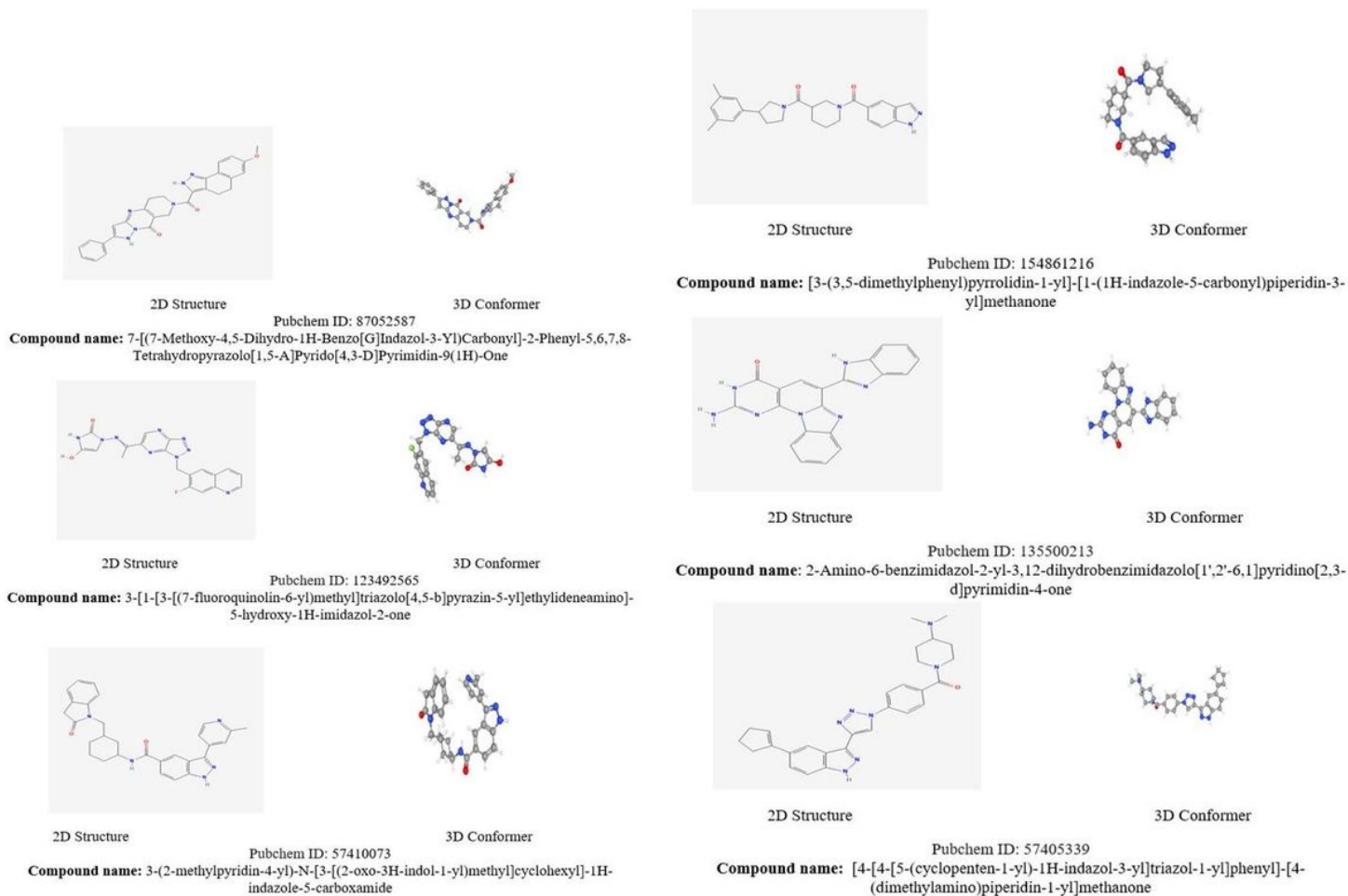


Figure 5

2D Structure and 3D Conformer of the six best hits

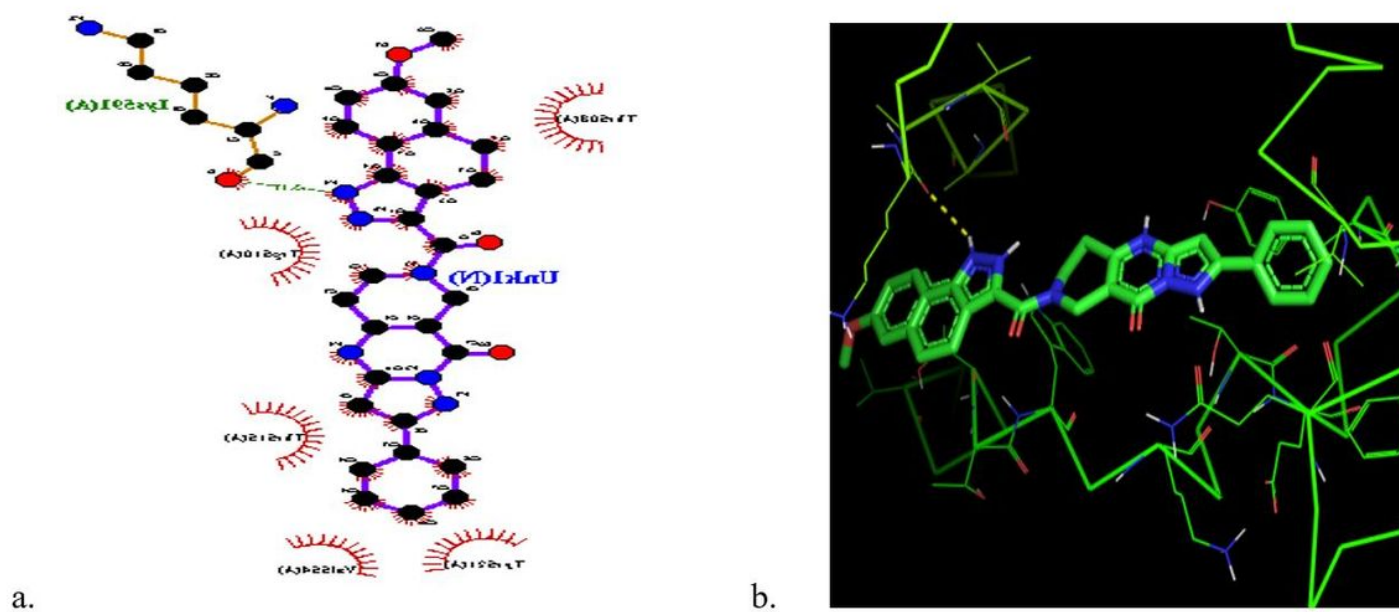


Figure 6

2D (a) and 3D (b) model interaction between best hit compound (ZINC97139187) and *PfAP2-I*.

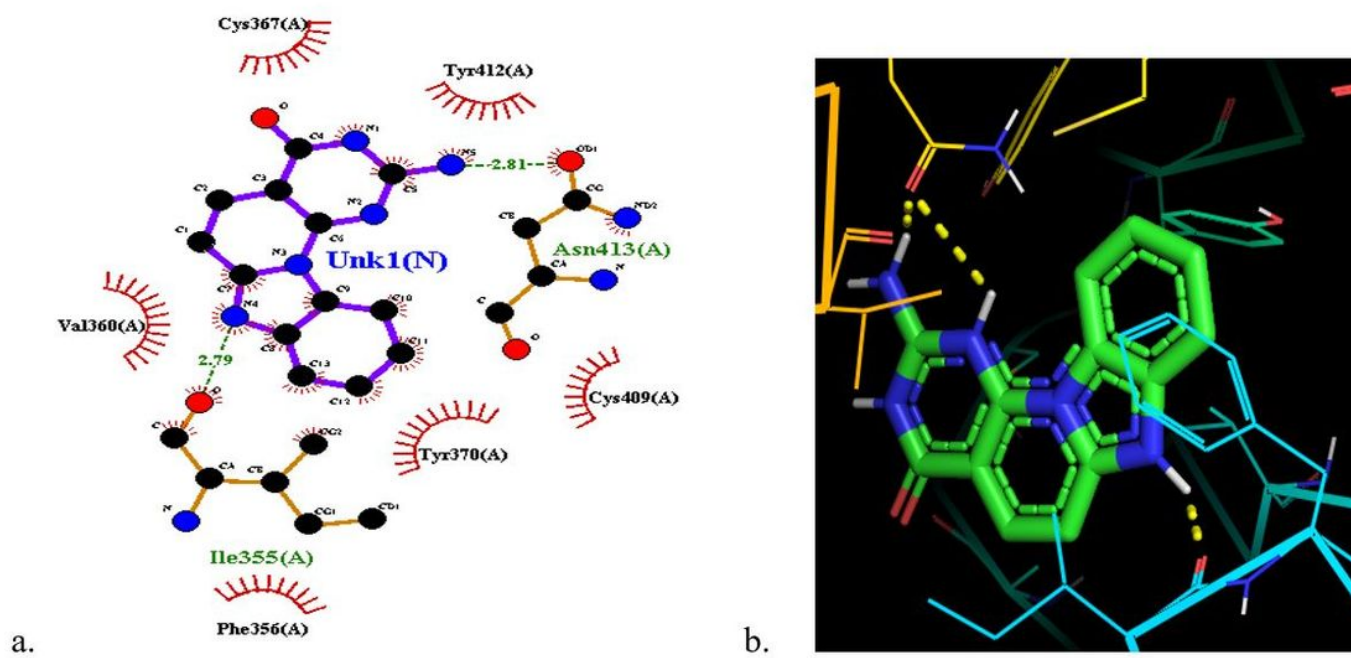
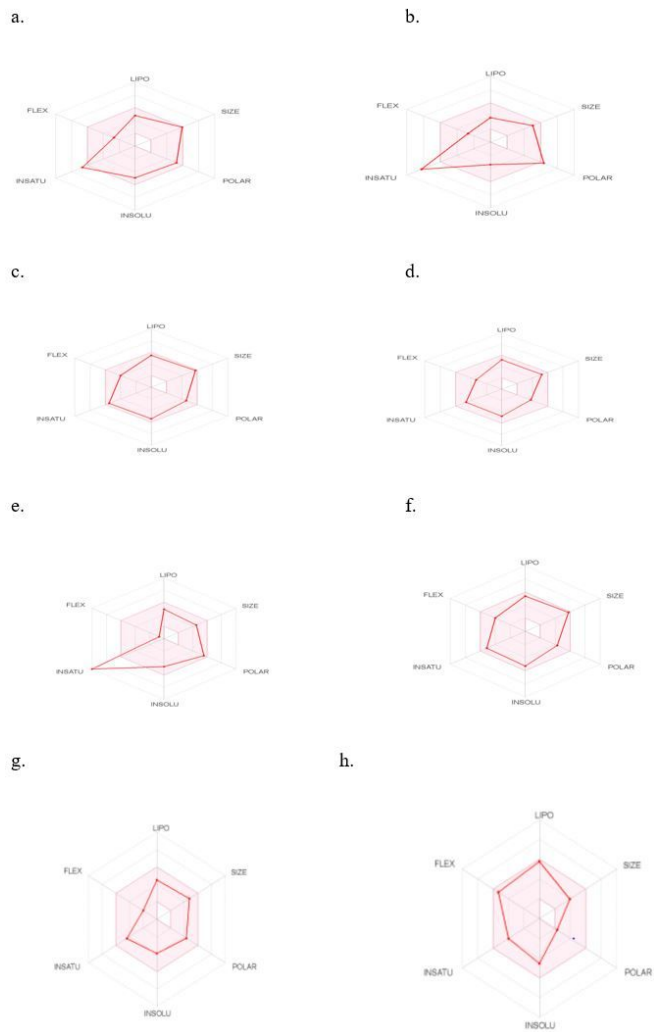


Figure 7

2D (a) and 3D (b) model interaction between best hit compound (MCULE-6567089130) and *PfBDP1*.



**Figure 8**

Oral bioavailability radar of the six compounds using SwissADME prediction