In silico Structure Prediction, Molecular Docking, and Dynamic Simulation of *Plasmodium falciparum* AP2-I Transcription Factor

David O Oladejo[1,2], Gbolahan O Oduselu[3], Titilope M Dokunmu[1,2], Itunuoluwa Isewon[1,4], Jelili Oyelade[1,4], Esther Okafor[2], Emeka EJ Iweala[1,2] and Ezekiel Adebiyi[1,4]

[1]Covenant Applied Informatics and Communication Africa Centre of Excellence (CApIC-ACE), Covenant University, Ota, Nigeria. [2]Department of Biochemistry, College of Science and Technology, Covenant University, Ota, Nigeria. [3]Department of Chemistry, College of Science and Technology, Covenant University, Ota, Nigeria. [4]Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Nigeria.

ABSTRACT: *Plasmodium falciparum* Apicomplexan Apetala 2 Invasion (*Pf*AP2-I) transcription factor (TF) is a protein that regulates the expression of a subset of gene families involved in *P. falciparum* red blood cell (RBC) invasion. Inhibiting *Pf*AP2-I TF with small molecules represents a potential new antimalarial therapeutic target to combat drug resistance, which this study aims to achieve. The 3D model structure of *Pf*AP2-I was predicted *ab initio* using ROBETTA prediction tool and was validated using Save server 6.0 and MolProbity. Computed Atlas of Surface Topography of proteins (CASTp) 3.0 was used to predict the active sites of the *Pf*AP2-I modeled structure. Pharmacophore modeling of the control ligand and *Pf*AP2-I modeled structure was carried out using the Pharmit server to obtain several compounds used for molecular docking analysis. Molecular docking and postdocking studies were conducted using AutoDock vina and Discovery studio. The designed ligands' toxicity predictions and *in silico* drug-likeness were performed using the SwissADME predictor and OSIRIS Property Explorer. The modeled protein structure from the ROBETTA showed a validation result of 96.827 for ERRAT, 90.2% of the amino acid residues in the most favored region for the Ramachandran plot, and MolProbity score of 1.30 in the 98th percentile. Five (5) best hit compounds from molecular docking analysis were selected based on their binding affinity (between –8.9 and –11.7 Kcal/mol) to the active site of *Pf*AP2-I and were considered for postdocking studies. For the absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties, compound MCULE-7146940834 had the highest drug score (0.63) and drug-likeness (6.76). MCULE-7146940834 maintained a stable conformation within the flexible protein's active site during simulation. The good, estimated binding energies, drug-likeness, drug score, and molecular dynamics simulation interaction observed for MCULE-7146940834 against *Pf*AP2-I show that MCULE-7146940834 can be considered a lead candidate for *Pf*AP2-I inhibition. Experimental validations should be carried out to ascertain the efficacy of these predicted best hit compounds.

KEYWORDS: *Pf*AP2-I, transcription factor, drug target, small molecules, molecular dynamics simulation, molecular docking

Background

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.[1] Many antimalarial drugs have been produced over the years, but resistance has been observed against most, including chloroquine, pyrimethamine, and proguanil.[2] 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.[3] 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.[4]

In silico techniques have been successful and have become powerful tools in the search to cure diseases,[5] 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 using sophisticated computers and information technology to speed up drug discovery, lead optimization, drug development, and design.[7] *In silico* methods such as molecular dynamics (MD) simulations, molecular docking, drug-likeness prediction, and ADMET (absorption, distribution, metabolism, elimination, and toxicity) studies are used to screen candidate drugs/molecules from various databases/libraries.[7] These *in silico* 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 compounds 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 "Sisyphean task," but recent work in drug discovery has shown the direct modulation of TF function by small molecules.[9-11] Transcription factors are proteins that bind to DNA sequences and control the stream of genetic information from DNA to mRNA.[12] Transcription factors, along with other proteins in a complex, control *P. 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 (RBCs).[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 site-based drugs.[11] One of such essential TFs found across the different *Plasmodium spp* is the Apetala 2—Invasion (AP2-I) TF (a member of the Apicomplexan Apetala 2 (ApiAP2) TF). *P. falciparum* Apetala 2—Invasion (*Pf*AP2-I) TF is a 183 kDa protein located on chromosome 10 of the *Plasmodium* genome whose primary function is to regulate RBC invasion genes.[15] *P. falciparum* Apetala 2—Invasion 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 3 AP2 domains, and only the third AP2 domain is essential for regulating a subset of genes involved in RBC Invasion[15,16] (Campbell et al[15]; *Pf*AP2-I is associated with several chromatin-associated proteins, including the *Plasmodium* bromodomain protein 1 (*Pf*BDP1)), and that complex formation is related to transcriptional regulation. *P. falciparum* Apetala 2—Invasion represents a potential new antimalarial therapeutic target as a critical regulator of RBC invasion.[17] It is essential to study and understand its functioning and determine drugs that can inhibit its activity for disrupting the parasite cycle in the human host and for designing effective therapies that can augment the efficacies of existing antimalarials. This study aims to identify small molecules with inhibitory potential against AP2-I regulatory action in *P. falciparum*.

Methods

PfAP2-I structure prediction

The structure of *Pf*AP2-I was modeled *ab initio* because there is a lack of experimentally validated structure for *Pf*AP2-I in the Protein Data Bank (PDB)[18] as well as the UniProt database (UniProtKB).[19] The protein ID of the target (*Pf*AP2-I 3D7 strain) was retrieved from the National Center for Biotechnology Information (NCBI) with the accession number *P*3D7_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 *Pf*AP2-I could not be done. *P. falciparum* Apetala 2—Invasion was then synthesized on

both the I-TASSER server (http://zhanglab.dcmb.med.umich. edu/I-TASSER)[21] and ROBETTA Baker server (http:// robetta.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 three-dimensional (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. The AlphaFold-predicted structure of *Pf*AP2-I is available on UniProt (www.beta.uniprot.org/uniprotkb/Q8IJW6/entry) and was compared with the predicted structure results from the I-TASSER server and ROBETTA Baker server.

Structure validation of modeled protein

PROCHECK[23,24] and ERRAT[25] on UCLA-DOE LAB—SAVES v6.0 was used to check for the quality of the modeled 3D structure of *Pf*AP2-I generated on the I-TASSER and ROBETTA Baker Laboratory. The .pdb file format of the modeled *Pf*AP2-I was uploaded on the UCLA-DOE LAB—SAVES v6.0 site to obtain the overall quality factor from ERRAT and Ramachandran plot statistics from PROCHECK. The overall quality factor of a protein structure 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.[26]

The Ramachandran plot is used to access a modeled protein's quality or an experimental structure. The Ramachandran plot statistics provide information on the total number of amino acid residues found in the favorable, allowed, and disallowed regions.[24] The prioritized *Pf*AP2-I 3D modeled structure from the ROBETTA Baker server and I-TASSER structure prediction and the AlphaFold-predicted structure of *Pf*AP2-I retrieved from UniProt were subjected to further structure validation using MolProbity[27] to determine the quality of prediction.

Active site prediction of AP2-I modeled structure

The active sites of modeled *Pf*AP2-I 3D7 structure were predicted using the Computed Atlas of Surface Topography of proteins (CASTp) 3.0[28] and *ConCavity*.[29] The 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.[30] *ConCavity* is an online service used for predicting protein-ligand-binding sites by combining evolutionary sequence conservation and 3D structure and works based on a confidentiality score (C-score). C-score is a confidence score of the predicted binding site. C-score values range from 0 to 1, where a higher score indicates a more reliable prediction. The modeled AP2-I 3D structure was submitted on the server in .pdb format. The necessary

amino acids for binding interactions predicted by the 2 servers were compared to determine the similarity between the 2 predicted active sites.

Pharmacophore-based virtual screening

Pharmacophore-based virtual screening was designed using Pharmit server.[31] 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.[32] The Pharmit server is based on a pharmacophore model using the AutoDock Vina scoring function.[33] A control ligand (3W7 from the COACH server) was selected for screening,[34] and both the modeled protein and control ligand were loaded into the Pharmit Server. The pharmacophore model was built using 6 features (Supplementary Table 1). The Pharmit filters hit screening for the pharmacophore modeling were set using the Lipinski rule of 5 and Veber's rule to minimize the results significantly and obtain the best possible inhibitors out of millions of drug-like compounds (Supplementary Table 1).

Protein and ligand preparation for molecular docking analysis

The modeled protein structure was defined as a receptor, while the complexed ligands were removed using Chimera software.[35] Furthermore, the protein was prepared by computing Gasteiger charges, adding polar hydrogens, and merging the nonpolar hydrogens using AutoDockTools 4.2.6.[36] In addition, OpenBabel software[37] was used to convert the .pdb files to the AutoDock docking format (.pdbqt), which was further used for the docking simulation.

Molecular docking analysis

Molecular docking of compounds against the active sites of *Pf*AP2-I was carried out using AutoDock Vina, an accessible graphical user interface (GUI) for the AutoDock 4.2 program.[38] The grid box was constructed using 80, 80, and 80 pointing in the x, y, and z directions, respectively, with a grid point spacing of 0.375 Å. The center grid box was set at −29.495, 57.365, and 45.252 Å around the amino acid residues in the active site of *Pf*AP2-I. Five (5) hits were then generated and ranked according to their binding affinities to verify the ligand-binding sites and postdocking analysis of the hit compounds was conducted using Discovery studio.[39]

In silico drug-likeness and toxicity predictions

The *in-silico* drug-likeness and toxicity predictions of the designed ligands were carried out using the SwissADME predictor[40] and OSIRIS Property Explorer.[41] SwissADME predictor provides information on the physicochemical properties, lipophilicity, water solubility, pharmacokinetics, drug-likeness, and medicinal chemistry of the compounds.[42] 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, drug score, as well as the mutagenic, tumorigenic, irritant, and reproductive risks.[43] Drug-likeness is a criterion for determining if a pharmacological substance possesses the characteristics of an orally active drug.[44] The Lipinski rule of 5 is an established concept based on drug-likeness. 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 5 H-bond donors, more than 10 H-bond acceptors, molecular weight must not be greater than 500, and the calculated LogP (cLogP) must not be greater than 5.[45] 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.[46]

Molecular dynamic simulation of the best hit compound

Molecular Operating Environment (MOE) 2019.01 software simulation module was used to carry out the molecular dynamic simulation of the best hit compound.[47,48] The protein and protein-ligand complex were protonated, the energy was minimized, and the AMBER 10: EHT force-field was parameterized at various times to get the stable conformer of the protein-ligand complex in an R-Field implicit solvation system.[48] The simulations required 3 steps. The initial step requires the molecular system to be heated to 310 K (37°C), followed by a 100-picosecond equilibration step at 310 K (37°C). The trajectory of the molecular system was then generated using the Nose–Poincare–Andersen (NPA) algorithm at 310 K for 1000 picoseconds (the time step of each simulation was set to 0.02 picoseconds). Visual molecular dynamics (VDM) software and Bio3D on the Galaxy Europe platform were used for visualizations and data processing. Principal components analysis (PCA) was used to model the system's key dynamics.[49] The simulation data set was reduced to a few essential components that define the directions with the most variance. The critical structural variants within the ensemble of protein structures were captured by ordering the principal components as eigenvectors based on the variance. The fraction of variance attributed to each principal component was visualized using an eigenvalue rank plot. The structural clustering based on the principal derived components, and residue-wise loadings were then used to determine how much each residue contributed to the first 2 principal components.

Results

PfAP2-I structure prediction

P. falciparum Apetala 2—Invasion structure prediction result from the I-TASSER server shows 4 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 are shown in Supplementary Table 2. *P. falciparum* Apetala 2—Invasion structure prediction result from the
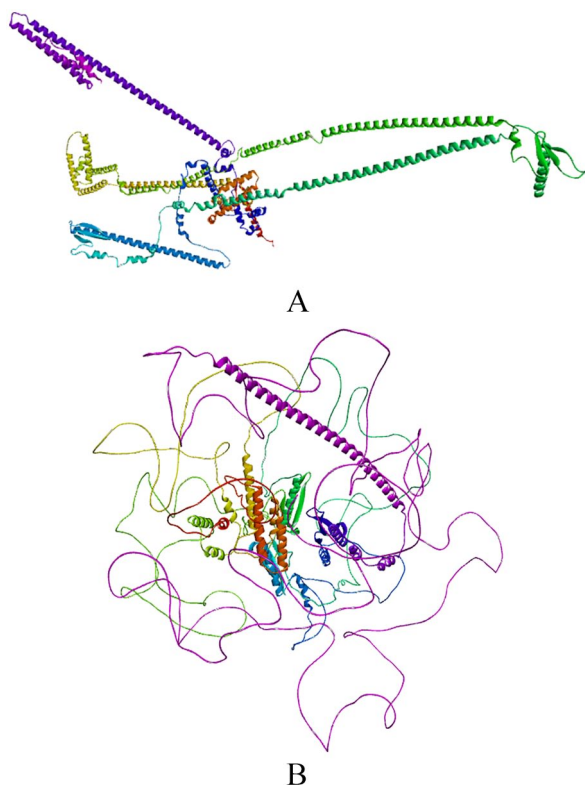

Figure 1. 3D modeled structure of *Pf*AP2-I TF from (A) ROBETTA Baker server and (B) AlphaFold-predicted structure of *Pf*AP2-I TF from UniProt. *Pf*AP2-I indicates *P. falciparum* Apicomplexan Apetala 2 Invasion; TF, transcription factor.

ROBETTA Baker server showed 5 predicted models using RoseTTAFold, and model 1 (1) (Figure 1A) was prioritized based on its highest C-score. The AlphaFold-predicted structure of *Pf*AP2-I was retrieved from UniProt (www.beta.uniprot.org/uniprotkb/Q8IJW6/entry) (Figure 1B). These 3 predicted structures from the I-TASSER server, ROBETTA Baker server, and AlphaFold were subjected to structure validation to determine the best-predicted structure.

Structure validation of modeled proteins

P. falciparum Apetala 2—Invasion 3D structure from the I-TASSER server and ROBETTA Baker server were validated using ERRAT and PROCHECK on UCLA-DOE LAB—SAVES v6.0. Model 1 from the ROBETTA Baker server showed the best result after validation. ERRAT value for model 1 from the ROBETTA Baker server was 96.827. Good resolution values are usually around 95% or higher (Figure 2). The Ramachandran plot statistics of model 1 from ROBETTA Baker server showed 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 (Figure 3). Model 1 from the ROBETTA Baker server and the AlphaFold-predicted structure of *Pf*AP2-I were subjected to further structure validation using MolProbity (www.molprobity.manchester.ac.uk). The predicted structure from the ROBETTA Baker server showed a better result than the AlphaFold-predicted structure of *Pf*AP2-I (Tables 1 and 2 and Supplementary Table 3).

Program: ERRAT2
File: model01.pdb
Chain#:A
Overall quality factor**: 96.827
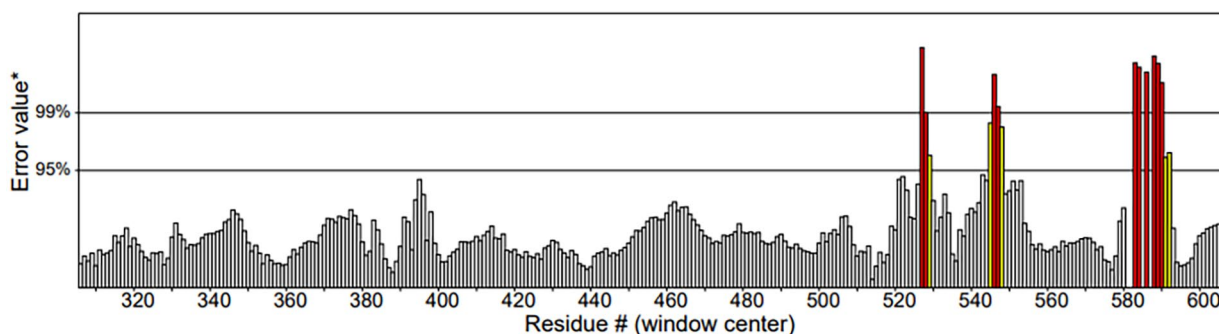
*On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

**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 *Pf*AP2-I-modeled structure. *Pf*AP2-I indicates *P. falciparum* Apicomplexan Apetala 2 Invasion.

Ramachandran Plot
saves

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
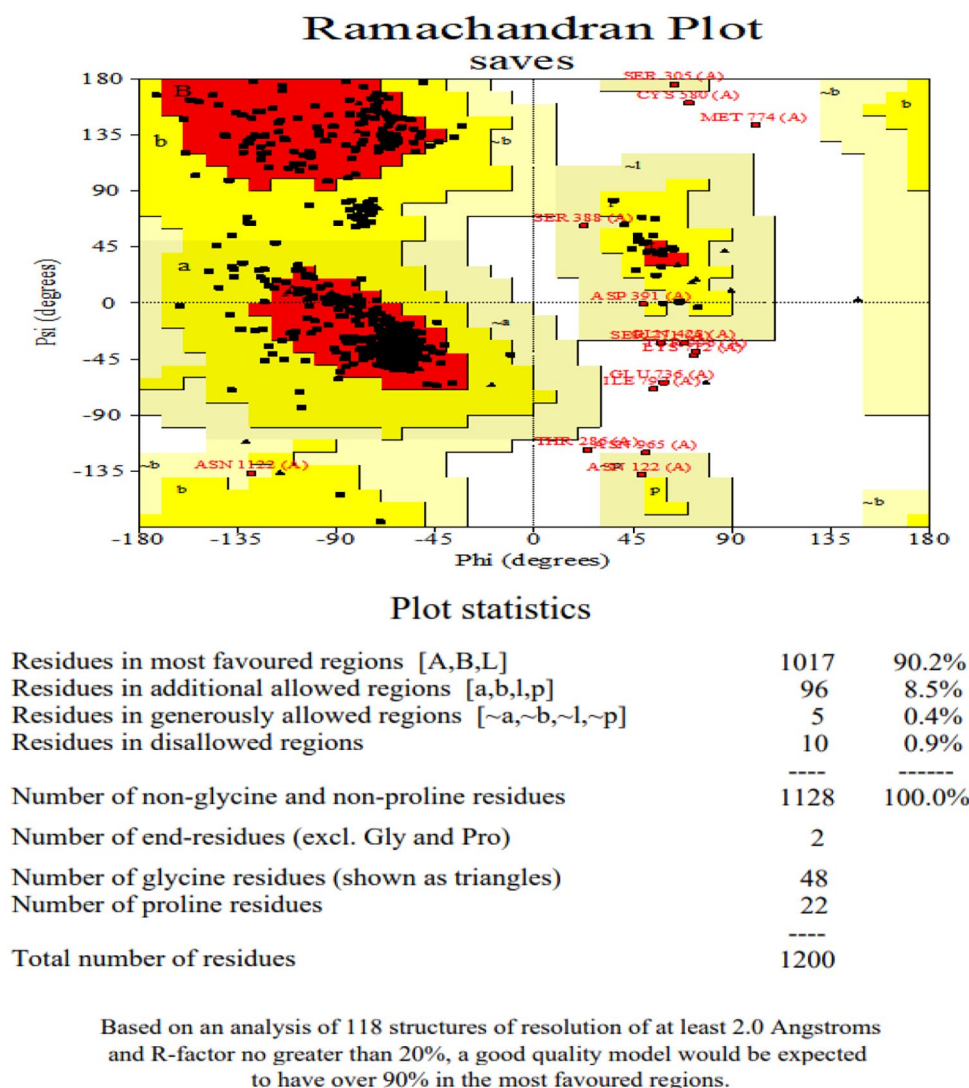to have over 90% in the most favoured regions.

Figure 3. Ramachandran plot statistic validation of *Pf*AP2-I-modeled structure.
*Pf*AP2-I indicates *P. falciparum* Apicomplexan Apetala 2 Invasion.

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 1 with an area (SA) of 15 407.119 and a volume (SA) of 13 411.050 (Figure 4) was selected as the preferred active site for the docking analysis. The amino acid residues in pocket ID 1 are listed in Supplementary Table 4.

Pharmacophore-based virtual screening and molecular docking analysis

Hit screening of drug-like compounds from 9 databases (Chemble, ChemDiv, ChemSpace, MCULE, MCULE-ULTIMATE, MolPort, NCI open repository, Lab Network, Zinc) on the Pharmit Server Engine gave a total of 8115 drug-like compounds (Table 3) that were used for the docking analysis including the reference compound (CHEMBL3359262) and the standard drug (chloroquine). Five (5) compounds were

identified based on the high-ranked auto-dock score from the docking analysis (Table 4). The chemical structure and IUPAC name of the 5 selected hit compounds against the *Pf*AP2-I active site are shown in Table 5.

Post-docking analysis

The 5 selected hit compounds from the molecular docking were subjected to postdocking analysis using Discovery studio. The 2D and 3D model interaction between the compounds and *Pf*AP2-I are shown in Figure 5.

In silico drug-likeness and toxicity predictions

Absorption, distribution, metabolism, elimination, and toxicity screening and toxicity testing results of the hit compounds including that of the reference compound (CHEMBL3359262) and standard drug (chloroquine) are shown in Tables 6 and 7, while their oral bioavailability radar are shown in Figure 6.

Table 1. Summary statistics of MolProbity structure validation analysis output for *Pf*AP2-I RoseTTAFold-predicted structure from Robetta Baker server.

All-atom contacts	Clashscore, all atoms:	1.93		99th percentile[a] (*N* = 1784, all resolutions)
	Clashscore is the number of serious steric overlaps (>0.4 Å) per 1000 atoms.			
Protein geometry	Poor rotamers	1	0.09%	Goal: <0.3%
	Favored rotamers	1099	99.55%	Goal: >98%
	Ramachandran outliers	14	1.17%	Goal: <0.05%
	Ramachandran favored	1139	95.08%	Goal: >98%
	Rama distribution *Z*-score	2.58 ± 0.24		Goal: abs (*Z* score) <2
	MolProbity score[b]	1.30		98th percentile[a] (N = 27 675, 0–99 Å)
	Cβ deviations >0.25 Å	1	0.09%	Goal: 0
	Bad bonds:	1/9809	0.01%	Goal: 0%
	Bad angles:	16/13 251	0.12%	Goal: <0.1%
Peptide omegas	Cis Prolines:	0/22	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis non-Prolines:	2/1177	0.17%	Goal: <0.05%
Low-resolution criteria	CaBLAM outliers	36	3.0%	Goal: <1.0%
	CA Geometry outliers	13	1.09%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/1382		
	Waters with clashes	0/0	0.00%	See UnDowser table for details

Abbreviation: *Pf*AP2-I, *P. falciparum* Apicomplexan Apetala 2 Invasion.
In the 2 column results, the left column gives the raw count, right column gives the percentage.
The green region signifies good, the yellow region signifies caution and the red region signifies warning.
[a]100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.
[b]MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

Table 2. Summary statistics of MolProbity structure validation analysis output for *Pf*AP2-I AlphaFold-predicted structure.

Protein geometry	Poor rotamers	102	6.90%	Goal: <0.3%
	Favored rotamers	1199	81.07%	Goal: >98%
	Ramachandran outliers	446	27.96%	Goal: <0.05%
	Ramachandran favored	872	54.67%	Goal: >98%
	Rama distribution *Z*-score	−6.42 ± 0.14		Goal: abs (*Z* score) <2
	Cβ deviations >0.25 Å	176	11.49%	Goal: 0
	Bad bonds:	1/12 934	0.01%	Goal: 0%
	Bad angles:	992/17 501	5.67%	Goal: <0.1%
Peptide omegas	Cis Prolines:	2/27	7.41%	Expected: ≤1 per chain, or ≤5%
	Cis non-Prolines:	194/1569	12.36%	Goal: <0.05%
	Twisted peptides:	526/1596	32.96%	Goal: 0
Low-resolution criteria	CaBLAM outliers	567	35.6%	Goal: <1.0%
	CA geometry outliers	660	41.43%	Goal: <0.5%
Additional validations	Tetrahedral geometry outliers	4		

Abbreviation: *Pf*AP2-I, *P. falciparum* Apicomplexan Apetala 2 Invasion.
In the 2 column results, the left column gives the raw count, and the right column gives the percentage.
The green region signifies good, the yellow region signifies caution and the red region signifies warning.
[a]100th percentile is the best among structures of comparable resolution; the 0th percentile is the worst.
[b]MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

Molecular dynamic simulation of the lead compound

The molecular dynamic simulation results of the lead compound MCULE-7146940834 are represented in Figures 7
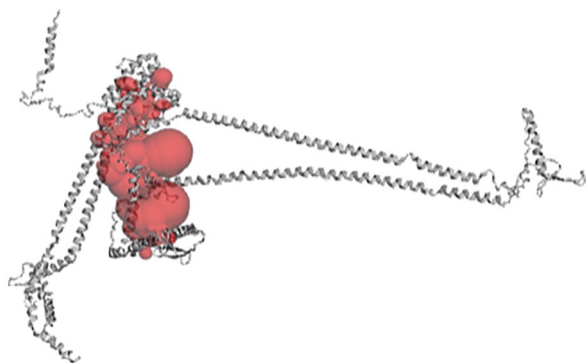

Figure 4. The binding pocket of 3D structure *Pf*AP2-I as computed by CASTp 3.0.
CASTp indicates Computed Atlas of Surface Topography of proteins; *Pf*AP2-I, *P. falciparum* Apicomplexan Apetala 2 Invasion.

and 8. The root mean square deviation (RMSD) plot, the RMSD histogram, and the Root Mean Square fluctuation (RMSF) plot are all shown in Figure 7. A stable fluctuation between 0.020 and 0.035 Å after an initial rise from 0.00 Å was observed for the RMSD (Figure 7A). Oscillations around positions 400, 920, and 1020 were observed for the RMSF (Figure 7B). The PCA plot, PCA cluster plot, and Principal components 1 (PC1) on RMSF are shown in Figure 8. PC2 versus PC1, PC2 versus PC3, and PC3 versus PC1 graphs, an eigenvalue rank plot (a and b), and the result of residue-wise loadings are included in the PCA plot and the PCA cluster plot. The cumulative variance is labeled for each data point in the eigenvalue plot. The first principal component (PC1) accounts for 8.75% of the overall variance, while the first 3 principal components account for 21.37% of the variance, according to the findings. Along the PC planes, a continuous color from blue to white to red was achieved (Figure 8A). Through the top 3 PC1, PC2, and PC3 spaces, the trajectory snapshots were divided into 2 different clusters of the hues black and red (Figure 8B). High peaks on residues 920 and

Table 3. Pharmacophore-based virtual screening of compounds from 9 databases on the Pharmit server.
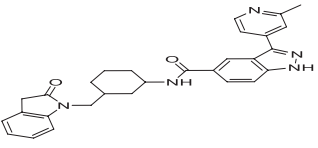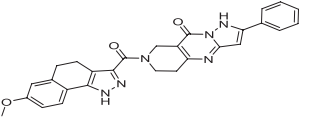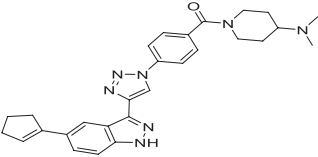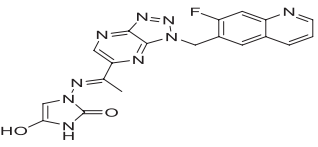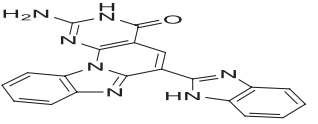
	PHARMIT DATABASE	CONFORMERS	MOLECULES	HITS
a.	CHEMBL25	23 136 925	1 752 844	182
b.	ChemDiv	21 562 497	1 456 120	77
c.	ChemSpace	250 205 463	50 181 678	1719
d.	MCULE	223 460 579	45 257 086	1716
e.	MCULE-ULTIMATE	378 880 344	126 471 502	2018
f.	MolPort	114 798 054	8 015 098	1534
g.	NCI Open Chemical Repository	574 117	52 237	–
h.	LabNetwork	22 051 020	1 794 286	138
i.	Zinc	123 399 574	13 190 317	1272
Total				8656

Table 4. Five selected hit compounds from the Pharmit server ranked based on the auto-dock score against *Pf*AP2-I active site.

DOCKED LIGAND	PHARMIT IDS	PUBCHEM IDS	MOLECULAR FORMULA	BINDING AFFINITY (KCAL/MOL)
a.	57410073	57410073	$C_{29}H_{29}N_5O_2$	−11.7
b.	MCULE-7146940834	87052587	$C_{28}H_{24}N_6O_3$	−11.1
c.	57405339	57405339	$C_{28}H_{31}N_7O$	−10.3
d.	CHEMBL3923620	123492565	$C_{19}H_{14}FN_9O_2$	−8.9
e.	MCULE-6567089130	135500213	$C_{20}H_{13}N_7O$	−8.9
f.	Reference compound (CHEMBL3359262)	52934178	$C_{22}H_{25}N_5O_2$	−8.8
g.	Chloroquine	2719	$C_{18}H_{26}ClN_3$	−5.5

Abbreviation: *Pf*AP2-I, *P. falciparum* Apicomplexan Apetala 2 Invasion.

Table 5. The Pharmit ID, PubChem ID, Chemical structure, and Compound name of the 5 selected hit compounds against *Pf*AP2-I active site.

DOCKED LIGAND	PHARMIT IDS	PUBCHEM IDS	CHEMICAL STRUCTURE	COMPOUND NAME
a.	57410073	57410073		3-(2-methylpyridin-4-yl)-N-[3-[(2-oxo-3H-indol-1-yl)methyl]cyclohexyl]-1H-indazole-5-carboxamide
b.	MCULE-7146940834	87052587		7-[(7-Methoxy-4,5-Dihydro-1H-Benzo[G]Indazol-3-Yl)Carbonyl]-2-Phenyl-5,6,7,8-Tetrahydropyrazolo[1,5-A]Pyrido[4,3-D]Pyrimidin-9(1H)-One
c.	57405339	57405339		[4-[4-[5-(cyclopenten-1-yl)-1H-indazol-3-yl]triazol-1-yl]phenyl]-[4-(dimethylamino)piperidin-1-yl]methanone
d.	CHEMBL3923620	123492565		3-[1-[3-[(7-fluoroquinolin-6-yl)methyl]triazolo[4,5-b]pyrazin-5-yl]ethylideneamino]-5-hydroxy-1H-imidazol-2-one
e.	MCULE-6567089130	135500213		2-Amino-6-benzimidazol-2-yl-3,12-dihydrobenzimidazolo[1',2'-6,1]pyridino[2,3-d]pyrimidin-4-one

Abbreviation: *Pf*AP2-I, *P. falciparum* Apicomplexan Apetala 2 Invasion.

1020 can be seen in the results of residue-wise loadings for PC1 (black) and PC2 (blue) (Figure 8C).

Discussion

This study presents a predicted 3D structure of *Pf*AP2-I (Figure 1) and molecular docking analysis to determine small molecule inhibitors against *Pf*AP2-I TF. The result of the modeled protein structure from the ROBETTA server[22] was prioritized based on structure validation results of 96.827 for ERRAT (Figure 2), 90.2% of the amino acid residues in the most favored region for the Ramachandran plot statistics[24] (Figure 3), and MolProbity score of 1.30 in the 98th percentile as well as a clashscore of 1.93 in the 99th percentile[27] (Tables 1 and 2). Pocket ID 1 (Figure 4) was selected as the preferred active site for the docking analysis. Molecular docking analysis of 8656 compounds generated from pharmacophore-based virtual screening of drug-like compounds from 9 databases (Chemble, ChemDiv, ChemSpace, MCULE, MCULE-ULTIMATE, MolPort, NCI open repository, Lab Network, Zinc) on the Pharmit Server Engine was carried out to determine the inhibitory potential of small molecules against *Pf*AP2-I TF[33] (Table 3). These compounds were used to prepare the ligand library and screened against the prepared 3D modeled structure of *Pf*AP2-I. Five (5) drug-like compounds were identified based on the docking analysis's

highest-ranked auto-dock score. They had lower binding energies when compared to the reference compound (CHEMBL3359262) and the standard drug (chloroquine), with binding energies of −8.8 and −5.5 Kcal/mol, respectively (Table 4). The chemical structure and IUPAC name of the 5 selected hit compounds against *Pf*AP2-I active site are shown in Table 5. Postdocking analysis using discovery studio shows the 2D and 3D model interaction between the 5 selected hit compounds and *Pf*AP2-I active sites (Figure 5).

Absorption, distribution, metabolism, elimination, and toxicity screening of the 5 drug-like compounds was based on bioavailability radar parameters, 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 6 properties (lipophilicity, size, polarity, insolubility, saturation, and flexibility)[42] (Figure 6). 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 500 g/mol, polarity (TPSA) between 20 and 130 Å², solubility (log S) not more than 6, saturation not less than 0.25 and flexibility, not more than 9 rotatable bonds.[42] MCULE-7146940834, 57410073, and 57405339 fell within the physicochemical space for all 6 properties and can be said to be orally bioavailable. CHEMBL3923620 and
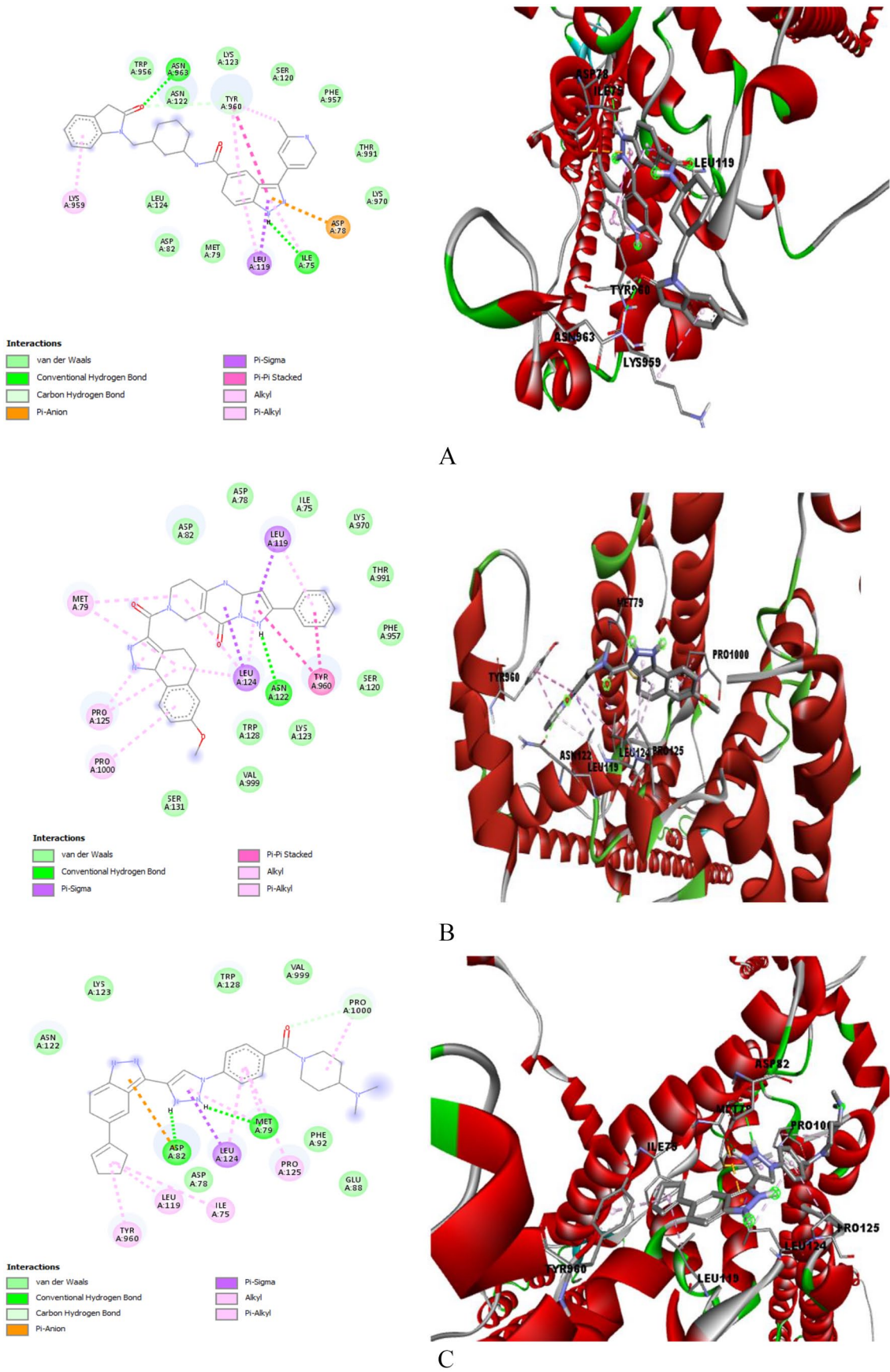
A

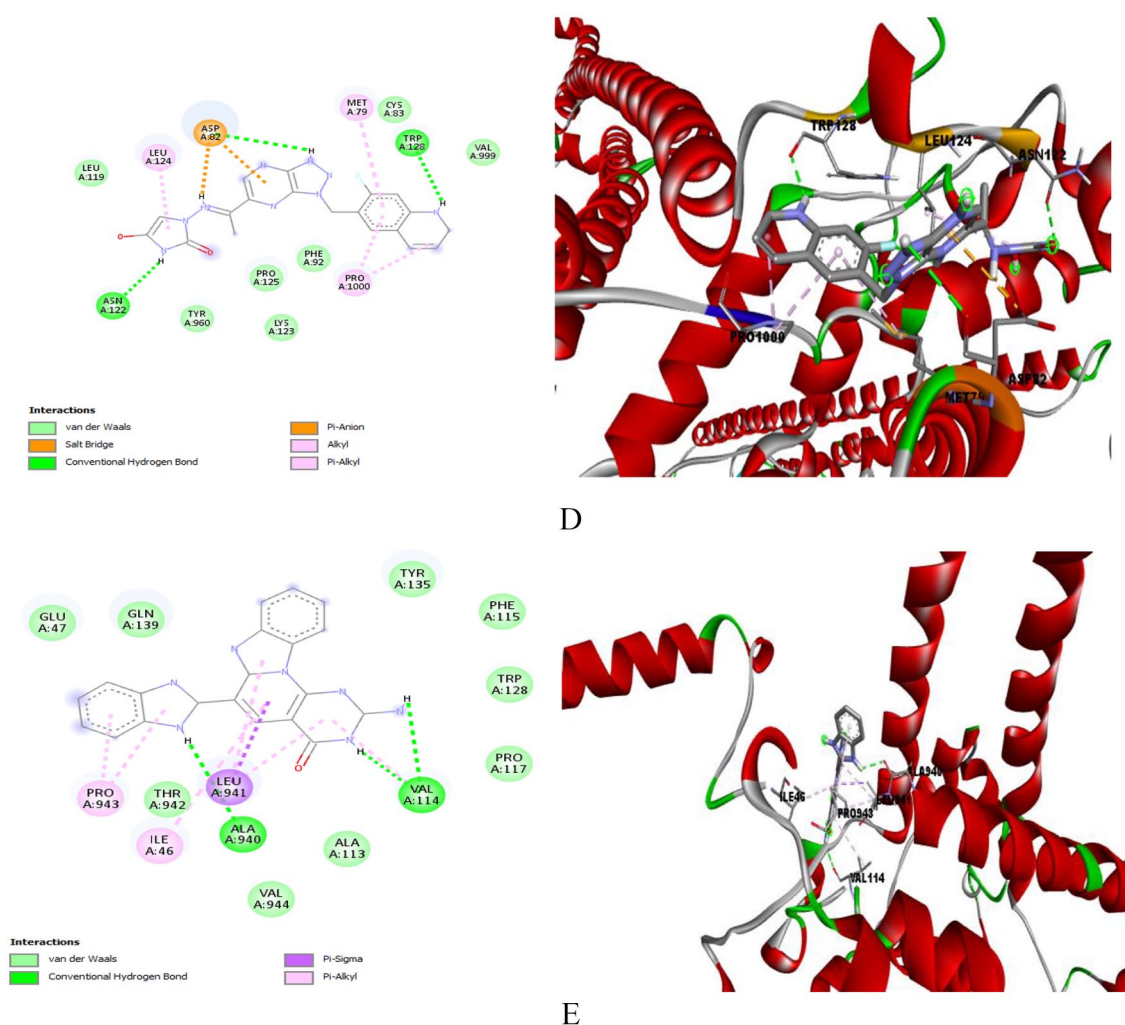

B

C

Figure 5. *(Continued)*

Figure 5. 2D and 3D model interactions between the 5 selected hit compounds and *Pf*AP2-I using Discovery studio. The nature of protein-ligand interactions is shown with different color legends: (A) 57410073, (B) MCULE-7146940834, (C) 57405339, (D) CHEMBL3923620, and (E) MCULE-6567089130.

*Pf*AP2-I indicates *P. falciparum* Apicomplexan Apetala 2 Invasion.

MCULE-6567089130 had a deviation in the degree of saturation outside the physicochemical space.

Tables 6 and 7 showed that all the compounds were within the minimum and maximum acceptable range. The 5 compounds had molecular weights of less than 500 g/mol, implying that when these compounds are administered as drugs, they are likely to be absorbed and reach the site of action. The numbers of rotatable bond (NRB), hydrogen bond acceptors (NHA), and hydrogen bond donors (NHD) in the 6 compounds (Table 6) follow Lipinski's rule of 5.[46]

Lipophilicity is an essential property in drug discovery as it is an indicator of pharmacodynamics, pharmacokinetics, and molecular toxicity.[50] 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 5 were found in all substances, including chloroquine, indicating good absorption, and penetration across cell membranes.[50]

The solubility of a molecule greatly facilitates major drug development activities, especially the ease of handling and

drug formulation. It is a significant property influencing the absorption and delivery of sufficient active ingredients in a small volume of pharmaceutical therapy.[51] Four compounds (MCULE-7146940834, 57410073, 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.[52] It is estimated that 50% to 90% of therapeutic molecules are substrates of 5 major isoforms (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4).[53] All 5 compounds have a high GI absorption and cannot cross the BBB. Four compounds are suitable inhibitors of CYP450 enzymes, while 2 are not (Table 6).

The selected compounds' physicochemical properties and toxicity risks were analyzed using *Osiris property explorer*.[54]

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

COMPOUNDS	PHYSICOCHEMICAL PROPERTIES			LIPOPHILICITY	WATER SOLUBILITY	PHARMACOKINETICS						MEDICINAL CHEMISTRY
	NRB	NHA	NHD	CLOGP	LOG	GI ABSORPTION	BBB PERMEANT	CYP1A2 INHIBITOR	CYP2C19 INHIBITOR	CYP2C9 INHIBITOR	CYP2D6 INHIBITOR	SYNTHETIC ACCESSIBILITY
a. 57410073	6	4	2	3.84	−5.42	High	No	No	Yes	Yes	Yes	4.14
b. MCULE-7146940834	4	5	2	1.44	−4.95	High	No	Yes	Yes	Yes	Yes	3.94
c. 57405339	6	5	1	2.76	−5.30	High	No	No	Yes	Yes	Yes	4.43
d. CHEMBL3923620	4	9	2	0.77	−3.41	High	No	No	No	No	No	3.50
e. MCULE-6567089130	1	4	3	2.11	−4.60	High	No	Yes	No	No	No	2.80
f. Reference compound (CHEMBL3359262)	3	4	1	2.57	−3.94	High	No	No	No	Yes	Yes	4.01
g. Chloroquine	1	2	8	4.01	−4.55	High	Yes	Yes	No	No	Yes	2.76

Abbreviations: ADME, absorption, distribution, metabolism, elimination; BBB, Blood-Brain Barrier; cLogP, consensus lipophilicity; GI, gastrointestinal; NHA, hydrogen bond acceptors; NHD, hydrogen bond donors; NRB, numbers of rotatable bond.

Table 7. OSIRIS property explorer prediction of physicochemical properties and toxicity risks of 6 compounds in comparison with chloroquine.

COMPOUNDS	PHYSICOCHEMICAL PROPERTIES					TOXICITY RISKS			
	MOLECULAR WEIGHT (G/MOL)	SOLUBILITY PREDICTION	TPSA (Å²)	DRUG-LIKENESS	DRUG SCORE	MUTAGENIC	TUMORIGENIC	IRRITANT	REPRODUCTIVE
a. 57410073	479.6	−6.08	90.98	3.86	0.26	Low Risk	High Risk	Low Risk	Low Risk
b. MCULE-7146940834	492.5	−4.3	108.38	6.76	0.63	Low Risk	Low Risk	Low Risk	Low Risk
c. 57405339	481.6	−3.71	82.94	−0.27	0.47	Low Risk	Low Risk	Low Risk	Low Risk
d. CHEMBL3923620	419.4	−6.48	139.76	−1.67	0.14	Medium risk	Low Risk	High 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. Reference compound (CHEMBL3359262)	391.0	−2.97	83.88	8.43	0.82	Low Risk	Low Risk	Low Risk	Low Risk
g. Chloroquine	319.9	−4.06	28.16	7.39	0.25	High Risk	Low Risk	High Risk	Low Risk

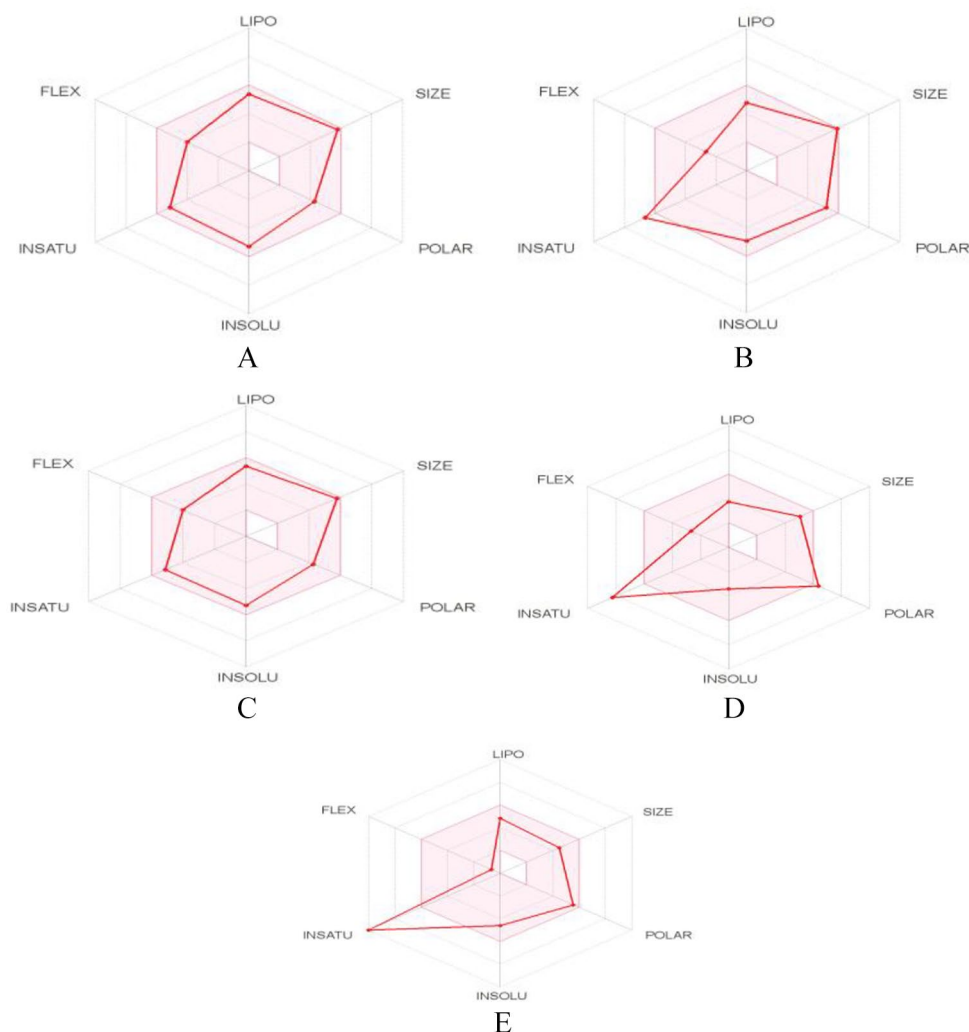Abbreviation: TPSA, total polar surface area.

Figure 6. Oral bioavailability radar of the 6 best hits vis-à-vis the reference compound and chloroquine using SwissADME prediction: (A) 57410073, (B) MCULE-7146940834, (C) 57405339, (D) CHEMBL3923620, and (E) MCULE-6567089130. The colored zone is the suitable physicochemical space for oral bioavailability. LIPO(Lipophilicity): $-0.7 < $ XLOGP3 < 5.0, SIZE: $150 \, g/mol < MW < 500 \, g/mol$, POLAR(Polarity): $20 \, Å^2 < TPSA < 130 \, Å^2$, INSOLU(Insolubility): $-6 < Log \, S \, (ESOL) < 0$, INSATU (Insaturation): $0.25 < $ Fraction Csp3 < 1, FLEX (Flexibility): $0 < $ Number of rotatable bonds < 9.
FLEX indicates flexibility; INSATU, insaturation; INSOLU, insolubility; LIPO, lipophilicity; MW, molecular weight; POLAR, polarity; TPSA, total polar surface area.

The drug-likeness, drug score, mutagenic tumorigenic, irritant, and reproductive properties were examined. Drug-likeness may be defined as a complex balance of molecular properties and structural features that determine whether a particular molecule is like the known drugs. A positive value for drug-likeness means that the molecule contains predominantly fragments frequently present in commercial drugs.[54] Three compounds, MCULE-7146940834, 57410073, and MCULE-6567089130, had positive values 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.[55] The higher the drug score, the better the chance to be a drug candidate. The drug score values 1.0, 0.8, and 0.6 are associated with no risk, medium risk, and high risk. Among the 6

compounds, MCULE-7146940834 had the highest drug score value of 0.63 (more elevated than chloroquine of 0.25), fell within the medium-risk range, and may be used as a drug molecule. The drug score values of 4 compounds (MCULE-7146940834, 57410073, MCULE-6567089130, and 57405339) with a range of 0.26 to 0.63 were more significant than that of chloroquine (0.25). Only one compound (CHEMBL3923620) was below that of chloroquine with a value of 0.14. CHEMBL3923620 possessed medium-risk mutagenic, low-risk tumorigenic, high-risk irritant, and low reproductive toxicity risks. Compound 57410073 showed low-risk mutagenic, high-risk tumorigenic, low-risk irritant, and low reproductive toxicity risks. All other compounds (MCULE-7146940834, 154861216, MCULE-6567089130, and 57405339) had low-risk mutagenic, low-risk tumorigenic, low-risk irritant, and low-risk reproductive toxicity risks.
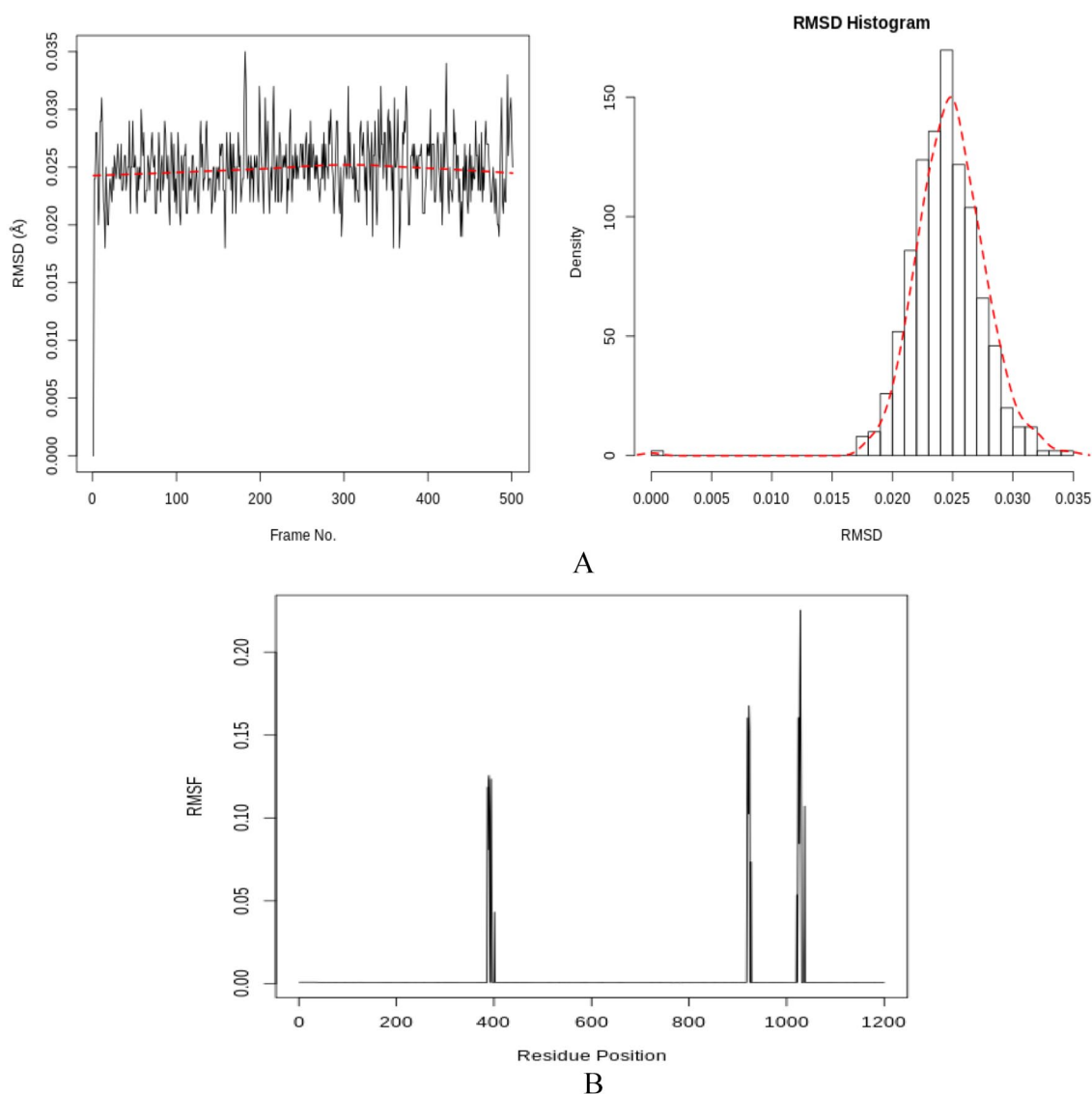
Figure 7. (A) RMSD time series and histogram for the interaction between MCULE-7146940834 and the active site of *Pf*AP2-I. (B) *Pf*AP2-I RMSF (Å) plot against the residue position.

*Pf*AP2-I indicates *P. falciparum* Apicomplexan Apetala 2 Invasion; RMSD, root mean square deviations; RMSF, root mean square fluctuation.

Molecular dynamics simulation was used to investigate the stability of the binding conformation and binding mode of the interactions between the lead compound, MCULE-7146940834 and *Pf*AP2-I active site residues.[56] Throughout the simulation period, the stability and conformation of the protein and ligand were checked using RMSD and RMSF. The PCA was used to determine the correlation between statistically meaningful conformations (significant global motions) collected during trajectory.[49] MCULE-7146940834 was stable with a single binding mode in the protein's active site, as indicated by the RMSD result (Figure 7A). The RMSD varies between 0.020 and 0.035, meaning that no substantial conformational changes occurred during the simulation phase, as evidenced by the histogram (Figure 7A). Significant shifts in the RMSF occur around positions 400, 920, and 1020, which correlate to the protein's surface's flexible loop sections (Figure 7B). The first 3 principal components account for 21.37% of

the overall variance, according to the eigenvalue rank plot (Figure 8A and B). The first principal component (PC1) accounts for 8.7% of the variance. The PCA analysis indicated a conformational shift in the protein backbone, which was grouped into 2 coordinate clusters in black (first cluster) and red (second cluster) using simple clustering in the principal component subspace. This is consistent with the results of residue-wise loadings (Figure 8C) and RMSF (Figure 7B), which showed significant changes around residues 920 and 1020. The ability of MCULE-7146940834 to achieve and maintain a stable conformation within the flexible protein's active site during simulation indicates the complex's stability, which is a crucial benefit for its inhibitory capability against *Pf*AP2-I.

Conclusion

The good, estimated binding energies, drug-likeness, and drug score values observed for the MCULE-7146940834 against
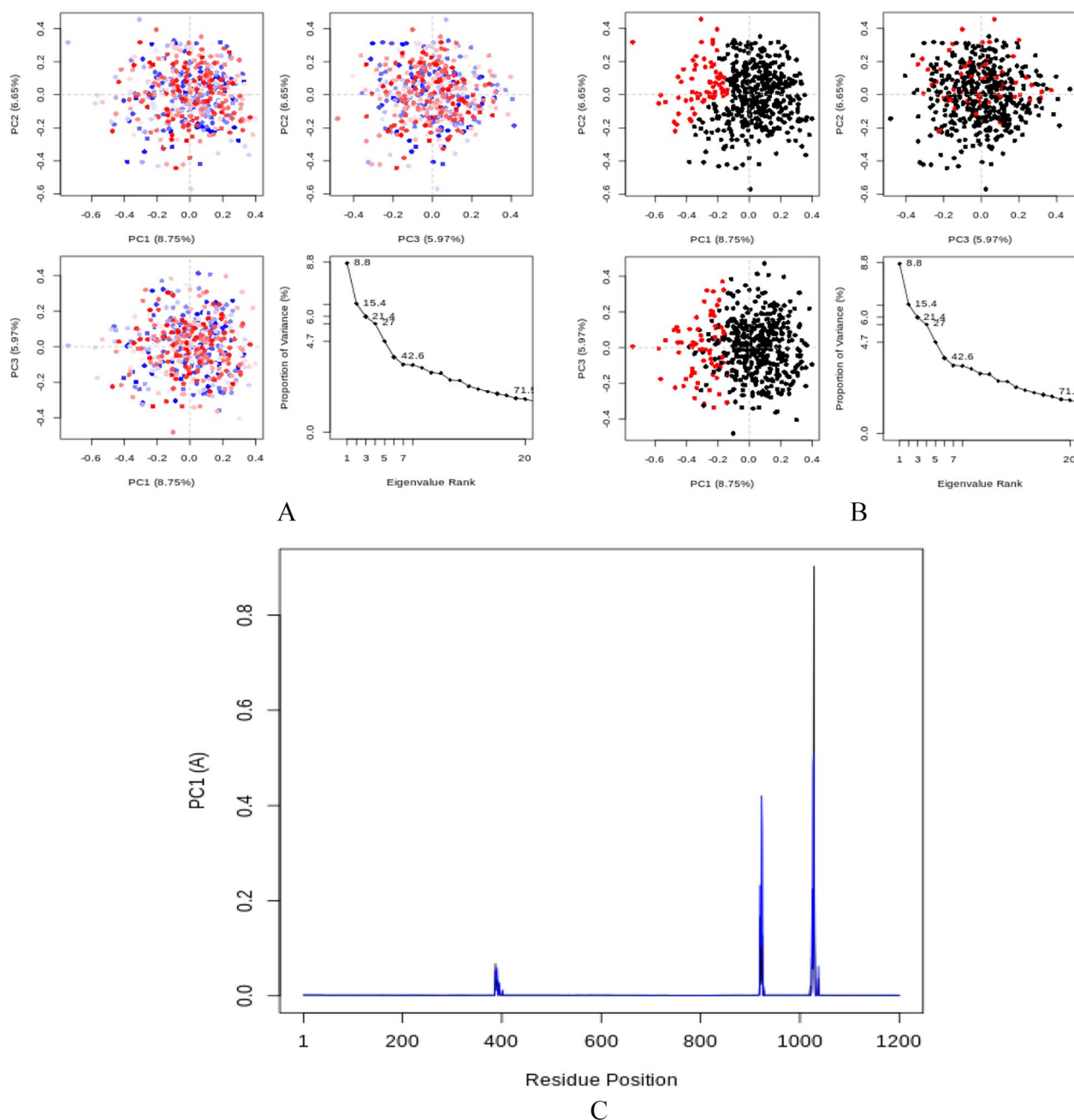
Figure 8. PCA results, comprising graphs of PC2 vs PC1, PC2 vs PC3, PC3 vs PC1, and an eigenvalue rank plot with the cumulative variance annotated for each data point. (A) PCA plots colored from blue to red in order of time; (B) PCA plots showing 2 different clusters colored black and red. (C) Residue-wise loadings for PC1 (black) and PC2 (blue).
PC indicates principal components 1; PCA, principal components analysis.

*Pf*AP2-I active site as well as the stable binding reliability by complex MD simulation from this study indicate that compound MCULE-7146940834 is a potential candidate for *Pf*AP2-I inhibition. Further preclinical experimental validations should be carried out to ascertain the efficacy of these predicted best hits.

Acknowledgements
Not applicable.

Author Contributions
All authors contributed significantly to the study: EA and DO design and led the study. GO, TD, II, EO, JO, and EEJI participated in the conduct and analysis, and EA, EEJI, GO, and TD participated in the final manuscript's conduct, review, and correction. All authors read and approved the final manuscript.

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.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Joseph Sahayarayan J, Soundar Rajan K, Nachiappan M, et al. Identification of potential drug target in malarial disease using molecular docking analysis. *Saudi J Biol Sci*. 2020;27:3327-3333.

2. Gregson A, Plowe CV. Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev*. 2005;57:117-145.

3. Nsanzabana C. Resistance to artemisinin combination therapies (ACTs): do not forget the partner drug! *Trop Med Infect Dis*. 2019;4:26. doi:10.3390/tropicalmed4010026

4. 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;10:470. doi:10.1007/S13205-020-02460-6

5. Sheikh I, Jiffri E, Ashraf G, Kamal M, Beg M. Structural studies on inhibitory mechanisms of antibiotic, corticosteroid and catecholamine molecules on lactoperoxidase. *Life Sci*. 2018;207:412-419.

6. Brogi S, Ramalho TC, Kuca K, Medina-Franco JL, Valko M. Editorial: in silico methods for drug design and discovery. *Front Chem*. 2020;8:612. doi:10.3389/FCHEM.2020.00612/FULL

7. Vardhan S, Sahoo SK. In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19. *Comput Biol Med*. 2020;124:103936.

8. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT. Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J*. 2016;14:177-184.

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;16:1030. doi:10.1186/S12864-015-2165-1

10. Papavassiliou KA, Papavassiliou AG. Transcription factor drug targets. *J Cell Biochem*. 2016;117:2693-2696. doi:10.1002/JCB.25605

11. Lambert M, Jambon S, Depauw S, David-Cordonnier MH. Targeting transcription factors for cancer treatment. *Molecules*. 2018;23:1479. doi:10.3390/molecules23061479

12. Kohzaki H, Murakami Y. Transcription factors and DNA replication origin selection. *BioEssays*. 2005;27:1107-1116. doi:10.1002/BIES.20316

13. Toenhake C, Bártfai R. What functional genomics has taught us about transcriptional regulation in malaria parasites. *Brief Funct Genomics*. 2019;18:290-301.

14. Toenhake CG. Transcription factors and gene regulatory elements in the human malaria parasite *Plasmodium falciparum*. https://repository.ubn.ru.nl/bitstream/handle/2066/215030/215030.pdf. Published 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 Pathog*. 2010;6:e1001165. doi:10.1371/JOURNAL.PPAT.1001165

16. Toenhake C, Fraschka S, Vijayabaskar M. Chromatin accessibility-based characterization of the gene regulatory network underlying *Plasmodium falciparum* blood-stage development. *Cell Host Microbe*. 2018;23:557-569.

17. Santos J, Josling G, Ross P, 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.

18. Burley S, Berman H, Bhikadiya C, Bi C. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res*. 2019;47:464-474.

19. Bateman A, Martin MJ, Orchard S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480-D489. doi:10.1093/NAR/GKAA1100

20. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:296-303.

21. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015;12:7-8.

22. Chivian D, Kim DE, Malmström L, et al. Automated prediction of CASP-5 structures using the Robetta server. *Proteins*. 2003;53:524-533. doi:10.1002/PROT.10529

23. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins*. 1992;12:345-364. doi:10.1002/PROT.340120407

24. Laskowski R, MacArthur M, Thornton J. PROCHECK: validation of protein-structure coordinates. In: *International Tables for Crystallography* (Vol. F: Crystallography of Biological Macromolecules). 2nd online ed. Wiley; 2012:684-687. https://onlinelibrary.wiley.com/iucr/itc/Fb/ch21o4v0001/ch21o4.pdf

25. Dym O, Eisenberg D, Yeates T. Detection of errors in protein models. In: Rossmann MG, Arnold E, eds. *International Tables for Crystallography*. Wiley; 2006:520-530.

26. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2:1511-1519. doi:10.1002/pro.5560020916

27. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27:293-315. doi:10.1002/pro.3330

28. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46:363-367.

29. 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:e1000585. doi:10.1371/JOURNAL.PCBI.1000585

30. 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. *Nucleic Acids Res*. 2006;34:W116-W118. doi:10.1093/NAR/GKL282

31. Sharma A, Yogavel M, Sharma A. Structural and functional attributes of malaria parasite diadenosine tetraphosphate hydrolase. *Sci Rep*. 2016;6:19981.

32. Wishart D, Feunang Y, Guo A, Lo E. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074-D1082.

33. Sunseri J, Koes D. Pharmit: interactive exploration of chemical space. *Nucleic Acids Res*. 2016;44:442-448.

34. Yang J, Roy A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 2013;29:2588-2595.

35. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-1612. doi:10.1002/jcc.20084

36. Huey R, Morris G, Forli S. *Using AutoDock 4 and AutoDock vina with AutoDockTools: A Tutorial*. The Scripps Research Institute Molecular Graphics Laboratory; 2012.

37. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform*. 2011;3:33. doi:10.1186/1758-2946-3-33

38. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30:2785-2791. doi:10.1002/jcc.21256

39. Biovia DS. *Discovery Studio Visualizer*. BIOVIA; 2017.

40. Lohidakshan 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. *Bangladesh J Pharmacol*. 2018;13:23-29.

41. Torres E, Moreno E, Ancizu S, et al. New 1, 4-di-N-oxide-quinoxaline-2-yl-methylene isonicotinic acid hydrazide derivatives as anti-Mycobacterium tuberculosis agents. *Bioorg Med Chem Lett*. 2011;21:3699-3703.

42. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7:42717.

43. Mahato S, Singh A, Rangan L, Jana C. 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. *Eur J Pharm Sci*. 2016;88:202-209.

44. Egbert M, Whitty A, Keserü GM, Vajda S. Why some targets benefit from beyond rule of five drugs. *J Med Chem*. 2019;62:10005-10025. doi:10.1021/ACS.JMEDCHEM.8B01732

45. Lipinski CA, Discovery M, Lombardo F, Lipinski CA, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23:3-25.

46. Behrouz S, Soltani Rad MN, Taghavi Shahraki B, Fathalipour M, Behrouz M, Mirkhani H. Design, synthesis, and in silico studies of novel eugenyloxy propanol azole derivatives having potent antinociceptive activity and evaluation of their β-adrenoceptor blocking property. *Mol Divers*. 2019;23:147-164. doi:10.1007/S11030-018-9867-7

47. Chemical Computing Group Inc. *Molecular Operating Environment*. Chemical Computing Group Inc.; 2016.

48. Johnson TO, Adegboyega AE, Ojo OA, et al. A computational approach to elucidate the interactions of chemicals from Artemisia annua targeted toward SARS-CoV-2 main protease inhibition for COVID-19 treatment. *Front Med (Lausanne)*. 2022;9:907583. doi:10.3389/FMED.2022.907583

49. Bray SA, Senapathi T, Barnett CB, Grüning BA. Intuitive, reproducible high-throughput molecular dynamics in Galaxy: a tutorial. *J Cheminform*. 2020;12:54. doi:10.1186/S13321-020-00451-6

50. Datta R, das D, das S. Efficient lipophilicity prediction of molecules employing deep-learning models. *Chemometr Intell Lab*. 2021;213:104309.

51. Wu CY, Benet LZ. Predicting drug disposition via application of BCS: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm Res*. 2005;22:11-23. doi:10.1007/s11095-004-9004-4

52. Savjani KT, Gajjar AK, Savjani JK. Drug solubility: importance and enhancement techniques. *ISRN Pharm*. 2012;2012:195727.

53. Di L. The role of drug metabolizing enzymes in clearance. *Expert Opin Drug Metab Toxicol*. 2014;10:379-393. doi:10.1517/17425255.2014.876006

54. Wolf CR, Smith G, Smith RL. Science, medicine, and the future: pharmacogenetics. *BMJ*. 2000;320:987-990.

55. Ursu O, Rayan A, Oprea T, Goldblum A, Oprea TI. Understanding drug-likeness. *Wiley Interdiscip Rev Comput Mol Sci*. 2011;1:760-781. doi:10.1002/wcms.52

56. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev*. 2006;26:531-568. doi:10.1002/med.20067