ISSN No:-2456-2165

Identification of Parasitic Cysteine Protease Inhibitors using Analog Design, Molecular Docking and Molecular Dynamics Studies

Ghodke Amol Yadavrao[1], Bhagwat Nivruttirao Poul[2], Chandrawadan Vishwambhar Panchal[3]
Suryawanshi Ashwini Ananda[4]
[1]Department of Pharmaceutical Sciences, Pacific Academy of Higher Education and Research University, Udaipur Rajasthan, India-313003
[2,3] Maharashtra College of Pharmacy, Nilanga, Maharashtra India-413521
[4] Shivai Charitable Trust's College of Pharmacy, Koregaonwadi, Omerga, Maharashtra India- 413606

Abstract:- In this study, *de novo* analogue design approach was adopted to find out novel chemical entities for the inhibition of cysteine protease falcipain-2 protein for *Plasmodium falciparum*. In this purpose a set 34 cysteine protease falcipain-2 protein inhibitors were collected from the literature and molecular docking performed. High binding energy complexes were used for analogue design in e-LEA3D: Chem Informatic Tools and Databases (http://chemoinfo.ipmc.cnrs.fr/eDESIGN/index.html). Best twelve analogues were selected from a number designed analogues based on dock score and binding interactions. Further, the drug likeness properties were analyzed and found that all selected analogues have potential being anti-malarial drug against cysteine protease falcipain-2. The root means square deviation (RMSD) and root means square fluctuation (RMSF) from the molecular dynamics simulation clearly indicated that selected analogues were formed stable complex with the cysteine protease falcipain-2. The binding energy of all selected analogues was calculated using MM-GBSA approach and found that each molecule shown strong affinity towards the receptor cavity. Therefore, the outcome of the study undoubtedly substantiated that final analogues have potential to be promising inhibitors for the cysteine protease falcipain-2 and can be used for therapeutic application in malaria-affected people subjected to experimental validation.

Keywords:- *Cysteine proteases, Plasmodium falciparum, De novo design, molecular docking, molecular dynamics.*

I. INTRODUCTION

Malaria, a mosquito-borne parasitic disease caused by most virulent human malaria parasite *Plasmodium falciparum*, one of the serious concern specially developing countries including India[1]. According to the World Health Organization (WHO) there were 216 million of malaria cases found in 2016, where as 445 000 deaths in the same year[2]. It is illustrated that Sub-Saharan African continues to carry an excessively high share of the global malaria burden. As per WHO, the Sub-Saharan Africa region was home to 90% of malaria cases and 92% of malaria deaths in 2015. Moreover Asia, and South America are also badly affected by this life threatening malaria disease[2]. As reported by the WHO, India alone contributing approximately 70% of the total cases of malaria evidenced in the South East Asian region[3]. Moreover, National Vector Borne Disease Control Programme (NVBDCP) reported that every year India affirmed approximately 1.5 million malaria cases, out of which 40–50% are because of Plasmodium falciparum. To overcome this deadly disease burden and treat, a number of drugs candidates are available in the market[4]. But due to the several knottiness in curing phase such as toxicity, high cost, and diminishing efficacy of drug molecules sometime treatments become very complicated and less effective. Recently, the artemisinin-based combination therapies (ACTs) are reported as highly effective against *P. falciparum* in Africa and most other malaria endemic countries[5-7]. Due to limited supply, highly expensive, and may suffer from resistance to both the artemisinin component and the partner drugs are limited to control the malaria[8]. Therefore, there is extensive need of new and improved drug candidates ideally directed against new protein or receptor targets to control the malaria globally. Among the several newly identified parasitic targets, cysteine proteases constitute is an important antimalarial chemotheraputic target[9]. A number of mammalian, viral, or protozoal cysteine proteases targets were used for the drug discovery approach including osteoporosis and Chagas disease[10]. Cysteine proteases mediates the protein hydrolysis with help of catalytic cysteine via nucleophilic attack on the carbonyl carbon of a susceptible peptide bond[11]. Cysteine proteases are categorized into clans and they are diverse in terms of sequence or structural identity and possibly arose independently[12]. Cys, His and Asn residues are used by Clan CA proteases and found always in the same order in the primary sequence of the cysteine protease. Clan CA belongs to the Family C1 (papain-family) cysteine proteases are well categorized for many eukaryotic organisms[13], and these are the best characterized cysteine proteases of Plasmodium. The clan CD, which exploits a catalytic His-Cys dyad and caspases in higher organisms and sequence analyses advise that members of the C13 and C14 families are present in plasmodia. Finally, the clan CE, which is considered as catalytic residues in the order of His, Glu (or Asp) and Cys, is also represented in the

P. falciparum genome. A number of reports have already been established the efficacy of peptidyl cysteine protease in terms of arresting and killing parasites in tissue culture models of parasite replication or cell invasion and have also confirmed their effectiveness in vivo[9]. Therefore, it is worth possible approach to develop the new and novel drugs for the treatment of malaria by targeting of parasite specific cysteine protease[14].

In the current work, an effort has inclined to design and identify new and novel chemical analogues for therapeutic application in malaria affected community using the pharma coinformatics approaches included *de novo* analogue design, molecular docking, *in silico* ADMET (absorption, distribution, metabolism, excretion and toxicity), molecular dynamics (MD) simulation and MM-GBSA (molecular mechanics - generalized Born and surface area) based binding energy calculation. A number of reported cysteine proteases inhibitors were collected form literature and docked inside the receptor cavity of cysteine proteases. The best docked complexes were used for analogues design. The final analogues were selected based on the dock score and number of binding interactions with catalytic amino residues of cysteine proteases. MD simulation study was performed on analogue-cysteine proteases complexes to check the stability in dynamic states. Finally, the binding affinity between analogues and cysteine proteases were explored by calculating the binding energy from the MD simulation trajectories.

II. EXPERIMENTAL

A. Dataset selection

A set of total 34 cysteine protease (for both falcipain-1, 2) inhibitors[15-18] for *Plasmodium falciparum* were selected for the molecular modeling and further study purpose. These inhibitors selection was made based on the inhibitory activities profile showed in similar type of bioassay studies against different parasites including malaria parasite, *Plasmodium falciparum.* SMILES (Simplified molecular-input line-entry system) notation and other molecular features details for selected 34 cysteine protease inhibitors are given in Table S1 (Supplementary file).

B. Target protein and ligands preparation for molecular docking

The high resolution crystal structure of cysteine protease falcipain-2 protein of *Plasmodium falciparum* was retrieved from Protein Data Bank (PDB ID: 3BPF)[19] available at www.rcsb.org, and selected as the bioactive target protein for molecular docking study. Using Auto Dock Tools v. 1.5.6, target protein structure (PDB ID: 3BPF) was prepared and saved as the input file (.pdbqt file format) required for docking analysis. During protein preparation step all water (H2O) molecules and bound ligand were removed from the macromolecular crystal structure. Polar hydrogen atoms were added to the protein structure and followed by Gasteiger charge was also computed and added. It was also checked for any missing atoms in the crystal protein structure and accordingly repaired for the same.

All selected 34 compounds were drawn using Chem Draw Ultra module of Chem Office 2002 software package and subsequently checked for any error in the ligands structure and resolved that immediately by cleaning before taking it to the further steps. Using the same AutoDock Tools v. 1.5.6, polar hydrogen atoms were added to the all ligand structures, and appropriate number of torsion were calculated and subsequently saved each ligand structure for further used.

C. Molecular Docking by Autodock Vina

Molecular docking simulations study for all 34 compounds was performed with Autodock Vina tool [20]. Docking protocol was performed by establishing the grid box centre measurement at (-51.45 Å) × (-11.066 Å) × (-15.676 Å) along the X, Y and Z axis, respectively, considering the protein active site dimensions within the specified measurement range. The box size dimensions for X, Y and Z coordinates were set at X = 80Å, Y = 80Å and Z = 80Å. The Autodock Vina configuration file was created with all these protein and ligands information used during docking procedure. Using the Lamarckian genetic algorithm (LGA) as scoring function the molecular docking was performed to generate the best binding poses between cysteine protease and each respective ligand.

D. De novo ligands design for cysteine protease inhibitors

The *de novo* ligand design was performed using e-LEA3D web server, which is the new version of the LEA3D engine or software tool [21-22] available at http://bioinfo.ipmc.cnrs.fr/lea.html. Based on the user defined docking functions, initially input was given such as coordinates of center of the binding site set as same which used previously during Autodock Vina docking procedure and other parameters such as binding site radius = 16 Å, weight in final score = 1. In the second step of the job submission in e-LEA3D web server, "*De novo* Drug Design" option was selected. Under default additional molecular parameters such as values for "conformational search by using the program Frog and maximum number of conformers" was set to 10, number of generations = 25 and population size = 40 defined and rest options left as default. The *de novo* ligand design web server e-LEA3D, utilizes a specific features of FDA approved drug fragments based approach for generation of new ligands or bioactive molecules. During *de novo* ligands design, PLANTS (Protein-Ligand ANT System), a semi-rigid docking program[23] was used for this study which generates user defined ten numbers of conformers or poses for each *de novo* designed ligand. PLANTS scoring function (*f*PLANTS) was used to compute the docking energy score for the *de novo* designed ligands and protein interactions. The following is the notation for the calculating scoring function[23].

ISSN No:-2456-2165

$$f\text{PLANTS} = f\,\text{plp} + f\,\text{clash} + f\,\text{tors} + c\,\text{site}$$

where, f plp = piecewise linear potential to model steric complementarity between ligand and protein atoms,

f clash = heavy atom potential to prevent internal ligand clashes f tors = ligand torsional potential, and

c site = distance-dependent quadratic potential to calculate the reference point of the ligand. PLANTS generated different binding poses were analyzed using PyMOL Molecular Graphic System tool for visualization of interactions map.

E. In silico ADMET, drug-likeness and synthetic accessibility analysis of de novo designed ligands

For computationally evaluating numerous important parameters necessary for consideration during drug developmental processes such as physicochemical and drug-likeness properties, nature of lipophilicity, water solubility, pharmacokinetics profiles and synthetic accessibility (SA). In the current work, the Swiss ADME [24] a web server based tool, developed and maintained by the SIB (Swiss Institute of Bioinformatics) was used. The OSIRIS Property Explorer a widely accepted open source tool (available at www.organic-chemistry.org/prog/peo/) was used for theoretically predicating toxicity risk assessment. The OSIRIS Property Explorer tool gives the predicated toxicity risks assessment results in provisions of high, medium and low toxic risk for any given compound.

F. Molecular dynamics simulations

In order to check dynamic information between protein and developed analogs the molecular dynamics (MD) simulation was performed. The dynamic behavior of the complexes can be explained by the simulating their internal motions or dynamic processes. In the current study, MD simulations study of protein-analog complexes was carried out using the Maestro-Desmond Interoperability Tools, in Schrödinger suite[25]. The systems of the complexes were prepared with help of the Desmond's 'System Builder' module. The orthorhombic shape of water box solvated by TIP3P (transferable intermolecular potential 3P) water model[26] was for each complex. The distance from the centre of the complex to the wall of the system was considered as 5Å. The OPLS_2005[27] molecular mechanics force field was used for all the systems. To neutralize the systems suitable number of Na^+/Cl^- ions were added after recalculating the charge of the systems. The each systems were minimized using the steepest descent (SD) and LBFGS (limited memory Broyden–Flecher–Goldfarb–Shanno) algorithm[28] with maximum 2000 iterations. The Nose– Hoover chain thermostat at 300 K and Martyna–Tobias–Klein barostat at 1.013 bar pressure was set for the simulation of the system. At each 10.0 ps of time interval the energy and trajectory frame of the simulated system was recorded. After system build up and minimization the MD simulations was performed for 40ns time span for each system in the NPT (number of atoms, pressure, area, and timescale) ensemble. The MD simulation was analyzed using the Simulation Interaction Diagram module of Maestro.

G. Binding free energy calculation using Prime Molecular Mechanics-Generalized Born and Surface Area (MM-GBSA) method

To calculate the free energy of binding the combined molecular mechanics energy and implicit solvation models are extensively used and applied in a number of studies[25, 29-30]. In the current work, complex between the protein and analogs were used to calculate binding free energies as the difference between the energy of the bound complex and the energy of the unbound target and analog as per following equation.

$$E\text{binding} \quad E\text{complex} \quad (E\text{protein} \quad E\text{ligand})$$

The energies of each protein and ligand were calculated individually after separation the calculation of energy of the complex form. All energies were calculated using OPLS_2005 force field[27] with the generalized Born implicit solvent model. Following equation was used to estimate the average binding free energy (ΔG).

$$G_{\text{binding}} \quad E_{\text{MM}} \quad G_{\text{GB}} \quad G_{\text{SA}} \quad T\ S$$

Where ΔEMM represents the bond, angle and dihedral energies (Δinternal), and van der Waals (ΔEvdw) energies; ΔGGB refers to is the electrostatic solvation energy (polar contribution), whereas and

ΔGSA is the non-electrostatic solvation component i.e. non-polar contribution. The Generalized Born (GB) is used to calculate the polar contribution and the non-polar energy is measured using the solvent accessible surface area (SASA)[31]. The –TΔS refers to the conformational entropy change is usually calculated by normal-mode analysis on a set of conformational snapshots obtained from MD simulations. in this work, from all 4000 frames, total 1000 frames with step size 4 were used to calculate the biding free energy.

III. RESULTS AND DISCUSSION

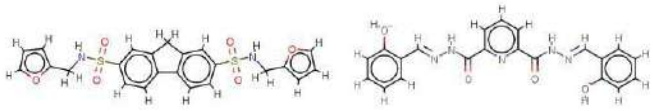
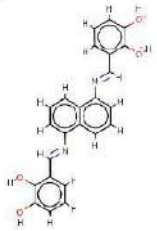
A. Analysis of molecular docking by Autodock Vina

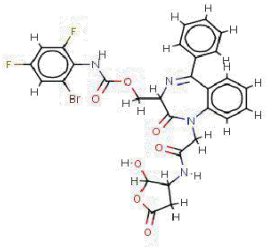
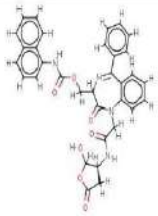
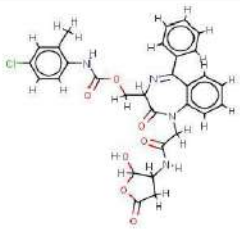
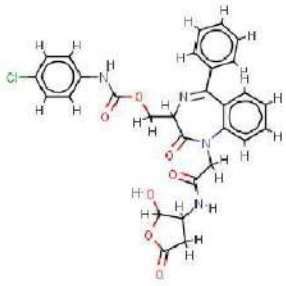
A set of 34 cysteine protease inhibitors[15-18] for *Plasmodium falciparum* (Cpd 1 to Cpd 34, Table S1, Supplementary file) were collected form literature subjected for *in silico* docking approach using Autodock Vina tool[20]. The best pose selection was made based on corresponding to the RMSD value of zero (0.0). In docking analysis the energy of binding or affinity values were obtained in best pose for all 34 compounds listed in Table S1 (Supplementary file). In order to select the best protein-ligand complexes for *de novo* designing, the binding energy cut of value was set as -8.0 Kcal/mol. Based on above cut of binding energy 11 compounds (Cpd 8, Cpd 14, Cpd 15, Cpd 24, Cpd 26, Cpd 27, Cpd 28, Cpd 29, Cpd 30, Cpd 31, Cpd 33) were found to be suitable for the *de novo* designing input. Subsequently, these

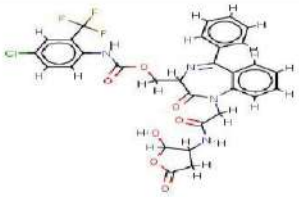
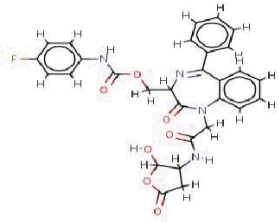
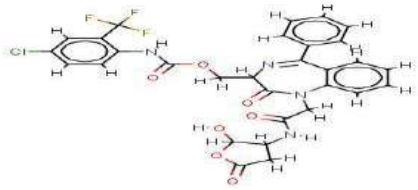
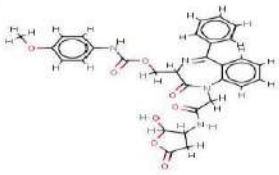
ISSN No:-2456-2165

compounds were analyzed for hydrogen bond (H-bond) interactions map. It was revealed that Cpd 26 has exhibited the highest binding affinity value of -9.5 kcal/mol and total of five H-bond interactions with the cysteine protease target protein. It was noted that, Cpd 26 has participated to form two H-bond interactions with amino acid residue Gly83, single H-bond interaction with Gln36, Asn173 and Trp206 residues of cysteine protease. The amino acid residue Gly83 believed to be key residue for implicating biological role which involvements was also observed in the co-crystallized ligand attached with PDB structure (PDB ID: 3BPF). This suggested that docking analysis was performed for this study purpose narrating the predicted accuracy of molecular docking supported by observing the incorporation of same or surrounding active site amino acid residues involvement in interacting radius. On the other hand, Cpd 8 and Cpd 28, both inhibitors have demonstrated to show lowest binding affinity value of -8.0 kcal/mol. The detailed of binding analysis and molecular interactions map of top 11 compounds in cysteine protease active site obtained through AutoDock Vina docking approach has been presented in Table 1.

Table 1. AutoDock Vina based molecular docking analysis of top 11 compounds having minimum binding affinity value of -8.0 kcal/mol with cysteine protease falcipain-2 protein of *Plasmodium falciparum* (PDB ID: 3BPF).

Compounds No.	Compounds Structure	Binding affinity (Kcal/mol)	Interacting residues in H-bond interactions	Total number of H-bond interactions
Cpd 8		-8.0	Trp 206 (Two H-bonds), Gln 209	3
Cpd 14		-8.8	Asn 173, His 174	2
Cpd 15		-8.2	Asp 35, Trp 206, Gln 209	3

			
Cpd 24	-8.3	Ala 157, Lys 160 (Three H-bonds)	4
			
Cpd 26	-9.5	Gln 36, Gly 83 (Two H-bonds), Asn 173, Trp 206	5
			
Cpd 27	-8.6	Gln 36, Gly 83, Trp 206	3
			
Cpd 28	-8.0	Gln 36, Gly 83 (Two H-bonds), Trp 206	4

			
Cpd 29	-9.3	Gln 36, Gly 83 (Two H-bonds), Trp 206	4
			
Cpd 30	-8.2	Asp 35 (Two H-bonds), Gln 36, Lys 37, Asn 173, Trp 206, Gln 209	7
			
Cpd 31	-8.3	Asp 35, Gln 36, Cys 39	3
			
Cpd 33	-8.1	Gln 36, Lys 37, Asn 173, Trp 206	4

ISSN No:-2456-2165

B. Analysis of receptor based de novo designed ligands inhibitors for cysteine protease

Initially, the *de novo* designed analogues were evaluated based on the empirical PLANTS score and molecular docking approach. Structurally diverse compounds were obtained from e-LEA3D server and ranked as per % of score observed. Further, the binding interactions between ligands and catalytic amino residues of cysteine protease were analyzed to select best analogues. Based on above analysis best 12 analogues (Analogue 1 - Analogue 12) were selected and retrieved from the *de novo* design program. Two dimensional (2D) molecular structures of 12 *de novo* designed analogues and their PLANTS based docked score are given in Figure 1. The uniqueness and novelty of the selected analogues were cross checked by executing a quick search on ChemSpider and PubChem chemical database with search options "Exact Match" and "Identical Structure", respectively for the both databases. No identical match was found for the 12 analogue in the search execution which undoubtedly explained that selected analogues are unique and novel.

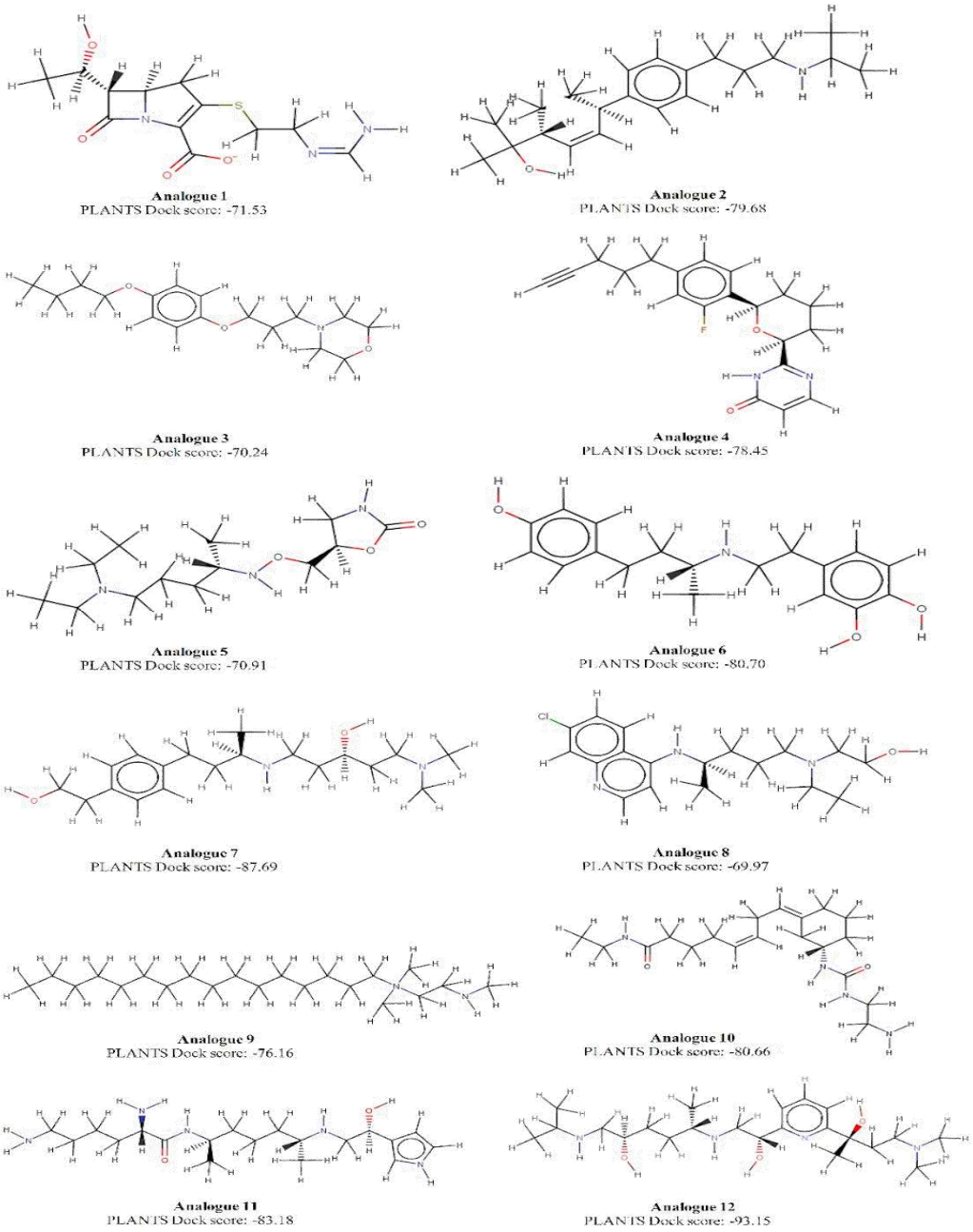


Fig 1:- 2D schematic molecular representation of 12 *de novo* designed analogues with their PLANTS based docked score.

ISSN No:-2456-2165

Table 2. PLANTS programme based docking analysis of top 12 *de novo* designed analogues with cysteine protease falcipain-2 protein of *Plasmodium falciparum* (PDB ID: 3BPF).

De novo designed compounds	SMILE structure format	Interacting residues in H- bond interactions	Number of H-bond interaction
Analogue 1	[O-]C(=O)C1=C(C[C@H]2N1C(=O)[C@@H]2[C@H](O)C) SCC/N=C\N	Asn 173 (Two H-bonds)	2
Analogue 2	CC(NCCCc1ccc(cc1)[C@@H](/C=C\[C@H](C(O)(C)C) C)C)C	Asp 234, His 174	2
Analogue 3	CCCCOc1ccc(cc1)OCCCN1CCOCC1	Ile 85	1
Analogue 4	C#CCCCc1ccc(c(c1)F)[C@H]1CCC[C@H](O1)c1nccc(=O)[nH]1	Asn 173 (Two H-bonds)	2
Analogue 5	CCN(CCC[C@@H](NOC[C@H]1CNC(=O)O1)C)CC	Asn 173, Trp 206	2
Analogue 6	C[C@H](CCc1ccc(cc1)O)NCCc1ccc(c(c1)O)O	Gly 83, Ile 85, His 174, Asp 234	4
Analogue 7	OCCc1ccc(cc1)CC[C@H](NCC[C@H](CCN(C)C)O)C	Gln 36, Gly 83 (Two H-bonds), Trp 206	4
Analogue8	OCCN(CCC[C@H](Nc1ccnc2c1ccc(c2)Cl)C)CC	Ile 85, Asn 173	2
Analogue9	CCCCCCCCCCCCCCCC[N+](CCNC)(C)C	Gln 36	1
Analogue10	NCCNC(=O)N[C@H]1CCC/C(=C/C/C=C\CCCC(=O)NC C)/C1	Gln 36, Ile 85, Asn 173 (Two H-bonds), His 174, Asp 234	6
Analogue11	NCCCC[C@@H](C(=O)N[C@@H](CCC[C@H](NC[C@H](c1 c[nH]cc1)O)C)C)N	Gln 36, Gly 83 (Two H-bonds), Asn 86, Asn 173	5
Analogue12	O[C@@H](CNC(C)C)CC[C@@H](NC[C@H](c1cccc(n1)[C@ @](CCN(C)C)(O)C)O)C	Gln 36, Gly 83 (Three H-bonds), Ile 85, His 174	6

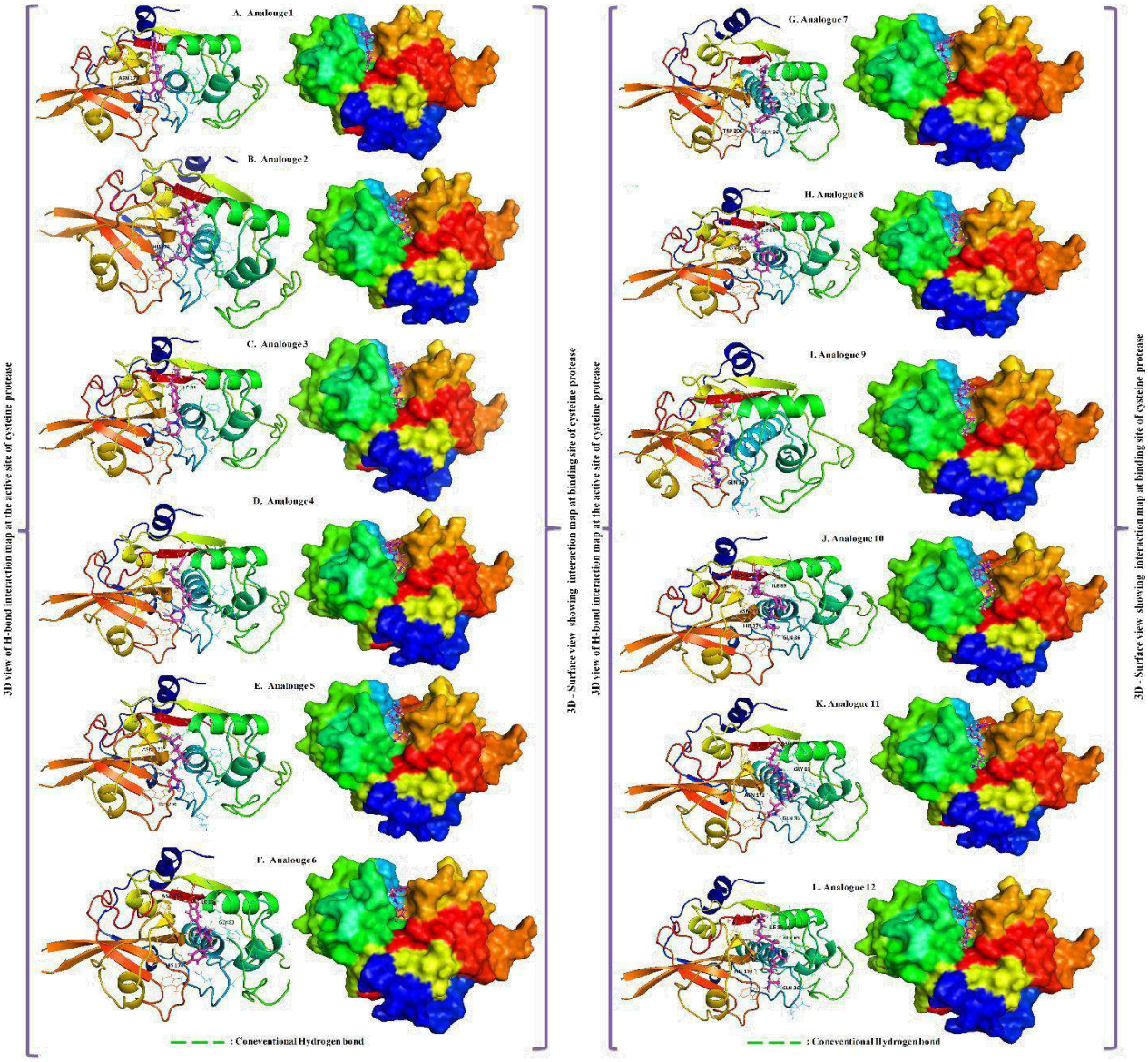


Fig 2:- Proposed PLANTS molecular docking based binding mode of *de novo* designed analogues (A: Analogue 1, B: Analogue 2, C: Analogue 3, D: Analogue 4, E: Analogue 5, and F: Analogue 6, G: Analogue 7, H: Analogue 8, I: Analogue 9, J: Analogue 10, K: Analogue 11, and L: Analogue 12) with cysteine protease of *Plasmodium falciparum*

C. In silico ADMET, drug-likeness and synthetic accessibility analysis of de novo designed ligands

The results of *in silico* analysis of different ADME (absorption, distribution, metabolism, and excretion) and physicochemical properties, drug-likeness profiles (Lipinski's rule of five (molecular weight = ≤ 500), H-bond acceptors (HBAs = ≤10), H-bond donors (HBDs = ≤5) and 1-octanol/water partition coefficient (logP = ≤5)), water solubility (Log*S* scale) and pharmacokinetics profiles (gastrointestinal absorption) were evaluated for all 12 de novo designed ligands and represented in Table 3. I was observed that except Analogue 1, all other analogues showed high gastrointestinal absorption rate indicating that majorly all designed analogue consists high affinity to be a druggable candidate for malaria treatment. In terms of agreement with 'Lipinski's rule of five' which was crucially measured as drug-likeness profiles for any ligands showed that ten numbers of analogues followed the rule, however, only two analogues (Analogue 2 and Analogue 11) violate the rule. Recommended values of other important physiochemical parameters such as number of rotatable bonds = ≤10, molar refractivity (MR) = 40 to 130, topological polar surface area (TPSA - Å2) ≤140 were also found to be under the acceptable range for most of the designed analogues. The above findings explained that all 12 selected analogues were shown to have drug likeness properties.

ISSN No:-2456-2165

Table 3. Lipinski's rule of five properties and other important physiochemical parameters of *de novo* designed analogues.

De novo designed analogues	Properties under Lipinski's rule of five				Other important physiochemical parameters			
	[1]MW(g/mol)	[2]HBA	[3]HBD	[4]Log*P*	[5]Rot. Bond	[6]MR	[7]TPSA	[8]vLOF
Analogue 1	298.34	5	2	0.28	6	77.08	144.35	0
Analogue 2	317.51	2	2	4.25	9	102.39	32.26	1
Analogue 3	293.40	4	0	1.83	9	88.47	30.93	0
Analogue 4	340.39	4	1	3.18	5	94.82	54.98	0
Analogue 5	273.37	5	2	0.88	10	77.66	62.83	0
Analogue 6	301.38	4	4	2.47	7	88.80	72.72	0
Analogue 7	322.49	4	3	2.16	12	97.27	55.73	0
Analogue 8	335.87	3	2	2.35	9	98.57	48.39	0
Analogue 9	327.61	1	1	1.01	18	108.24	12.03	0
Analogue 10	350.50	3	4	1.66	13	102.29	96.25	0
Analogue 11	367.53	5	6	-0.31	15	105.44	129.19	1
Analogue 12	410.59	7	5	-0.01	14	118.14	100.88	0

[1]Molecular weight; [2]Hydrogen bond acceptor; [3]Hydrogen bond donor; [4]Octanol water coefficient; [5]Number of rotatable bonds; [6]Molecular refractivity; [7]Topological polar surface area; [8]Violations of Lipinski's rule of five OSIRIS Property Explorer tool predicated the potential toxicity effects assessment for all the 12 *de novo* designed analogues and their results were exhibited as medium or low Toxicity risk in terms of mutagenicity or tumorigenicity. In Table 4, rigorously evaluated mutagenic or tumorigenic risk assessment characteristics has been represented for all the 12 *de novo* designed analogues. It was appreciated to notice the predication judgments that no indication of mutagenicity or tumorigenicity found for any designed analogue.

ISSN No:-2456-2165

Table 4. ADME and different pharmacological properties, drug-likeness and medicinal chemistry profiles predicted using SwissADME and Osiris property explorer tool for 12 *de novo* designed analogues.

De novo designed analogues		Analogue 1	Analogue 2	Analogue 3	Analogue 4	Analogue 5	Analogue 6	Analogue 7	Analogue 8	Analogue 9	Analogue 10	Analogue 11	Analogue 12
Water solubility	Log *S*	VS	MS	S	S	VS	S	S	S	MS	VS	VS	VS
Pharmacokinetc profiles	GI absorption	Low	High	High	High	High	High	High	High	High	High	High	High
	log *K*p (cm/s)	-8.62	-4.9	-5.86	-6.22	-7.03	-5.7	-6.67	-5.81	-2.9	-7.72	-8.58	-8.61
	CYP1A2 inhibitor	No	Yes	No	No	No	No	No	Yes	No	No	No	No
	CYP2C19 inhibitor	No	No	No	Yes	No	No	No	No	No	No	No	No
	CYP2C9 inhibitor	No	No	No	Yes	No	No	No	No	No	No	No	No
	CYP2D6 inhibitor	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	No
	CYP3A4 inhibitor	No	Yes	No	No	No	No	No	Yes	No	No	No	No
Drug-likeness profiles	Bioavailability score		0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.55
Medicinal chemistry profiles	Synthetic accessibility	4.46	3.55	2.42	4.01	4.15	2.43	3.05	2.82	2.92	4.05	4	4.71
	PAINS alerts	0	0	0	0	0	1	0	0	0	0	0	0
Toxicity risk assessment	Mutagenicity risk	Low	Low	Low	Low	Low	Low	Med	Med	Low	Low	Low	Low
	Tumorigenicity risk	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low

De novo designs process cannot give the approval that all designed analogues or compounds are feasible for chemical synthesis. As per reports, *de novo* designed molecules requires synthetic feasibility study of the compound for implementing experimental validation of its biological activity. It is always important to consider those compounds which can be synthesized with reasonable effort. Also in many areas of the drug discovery process and medicinal chemistry research, synthetic accessibility (SA) is considered as a major endpoint to measure the potentiality of a compound to estimate ease of synthesis of drug-like molecules is needed. Consequently selected analogues were considered for SA analysis. In this purpose, the SwissADME web server was used to evaluate SA values for all *de novo* designed analogues and are listed in Table 4. The recommended scoring scale for the predicted SA value ranges from '1' (very easy to synthesize) to '10' (very challenging to synthesize). From Table 4 it was observed that SA value of all *de novo* designed analogues in the range of 2.42 to 4.71, which indicates that not a single analogue is difficult to synthesize. However, SA values of Analogue 1, Analogue 6, Analogue 8, Analogue 10, and Analogue 12 obtained as 4.46, 2.43, 2.82, 4.05 and 4.71, respectively.

D. Molecular dynamics

ISSN No:-2456-2165

The best docked pose of each of 12 selected analogues and Cpd 15 (most active compound) were subjected to 40ns time span of MD simulation study to ensure the protein-ligand stability. The RMSD and RMSF data of all complexes were analyzed. The average RMSD of the complexes were found to be 1.648, 1.880, 2.032, 1.950, 2.237, 2.115, 2.049, 1.916, 1.805, 1.945, 1.916, 1.991 and 1.899Å for Cpd 15, Analogue 1, Analogue 2, Analogue 3, Analogue 4, Analogue 5, Analogue 6, Analogue 7, Analogue 8, Analogue 9, Analogue 10, Analogue 11 and Analogue 12 respectively. The plot of RMSD vs time is given in Figure 2. In detailed, it was observed that all systems initially fluctuated but attained the consistency between 1.50 to 2.50Å except Analogue 11. The Figure 3 explained that RMSD values of Analogue 11 were fluctuated much during entire simulation time span which suggested the somewhat instability of the protein-ligand complex.

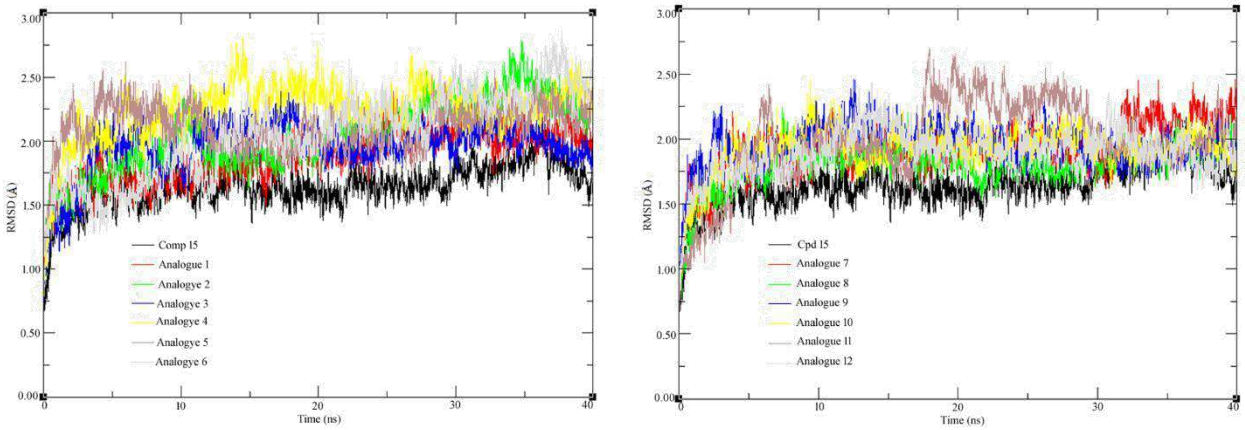


Fig 3:- RMSD vs Time of Cpd 15 and *de novo* designed 12 analogues

In order to explore the fluctuation of individual amino residues the RMSF values from 40ns MD simulation trajectories were analysed. The RMSF of a particular residue in the MD trajectories was estimated by calculating the average of all the atoms in that residue. In the current work, difference between highest and lowest RMSF values were found to be 3.636, 3.305, 4.328, 4.195, 3.704, 3.494, 4.317, 4517, 3.741, 3.831, 3.589, 3.690 and 4.107Å for Cpd 15, Analogue 1, Analogue 2, Analogue 3, Analogue 4, Analogue 5, Analogue 6, Analogue 7, Analogue 8,

Analogue 9, Analogue 10, Analogue 11 and Analogue 12 respectively. The RMSF values with respect to the residue index are plotted in the Figure 4. The Figure 4 indicated that amino residues around Val187, Asn188 and Gly194 were fluctuated up to 4Å for the all systems. Moreover, Analogue 1 and Analogue 6 were showed high RMSF values around from Leu113 to Thr120 about range of 2.331 to 3.218Å. Analogue 2, Analogue 3, Analogue 4, Analogue 6, Analogue 10 and Analogue 11 were found to be high RMSF values region from Gly79 to Gly82. The amino residues Glu13 to Phe17 region of the protein molecule showed higher RMSF valued in case of Analogue 3, Analogue 5 and Analogue 9. The amino residues His27 and Gly168 were found to be higher fluctuation for Analogue 2 and Analogue 1 respectively. In case of Cpd 15, amino acid region Gly29 to Thr31 were showed high fluctuation to about 3.608Å. The possible reason of fluctuation of these region may be a lack of interactions between amino residues and ligands in these regions.

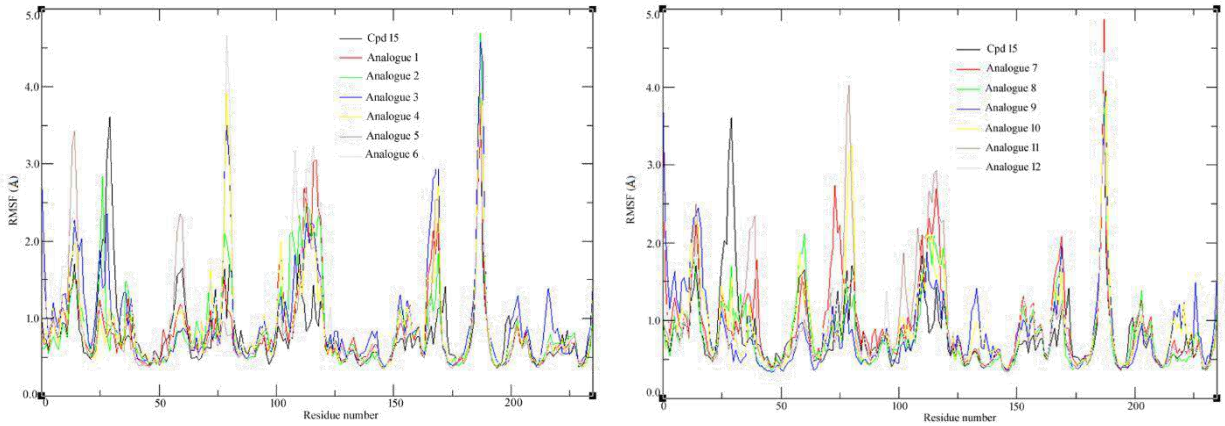


Fig 4:- RMSF vs residue number of Cpd 15 and *de novo* designed 12 analogues

E. Free energy calculation using MM-GBSA

The binding affinity of all twelve analogs and most active compound towards the cysteine protease protein were explored using the binding energy calculation based on MM-GBSA method. The *ΔGbind* was calculated from 1000 frames selected from 40ns MD trajectory of each system using step_size 4. The high negative *ΔGbind* value explain more affinity towards the receptor cavity. The average binding energy of complexes with final analogs and Cpd 15 were found to be -51.012, - 49.062, -40.954, -31.289, -34.594, -45.374, -46.74, -46.926, -35.962, -46.738, -39.784, -52.384, -38.185 and -50.010 kcal/mol for Cpd 15, Analogue 1, Analogue 2, Analogue 3, Analogue 4, Analogue 5, Analogue 6, Analogue 7, Analogue 8, Analogue 9, Analogue 10, Analogue 11 and Analogue 12 respectively. The binding energy trajectory with respect to the number of frames are given in Figure 5. It can be observed that Analogue 10 gives higher binding energy compare to most active compound. Moreover, Analogue 1, Analogue 6, Analogue 8 and Analogue 12 were found to have comparable binding affinity towards the receptor with most active compound. Rest of the analogs were found to have low binding affinity compare to Cpd 15 compound.

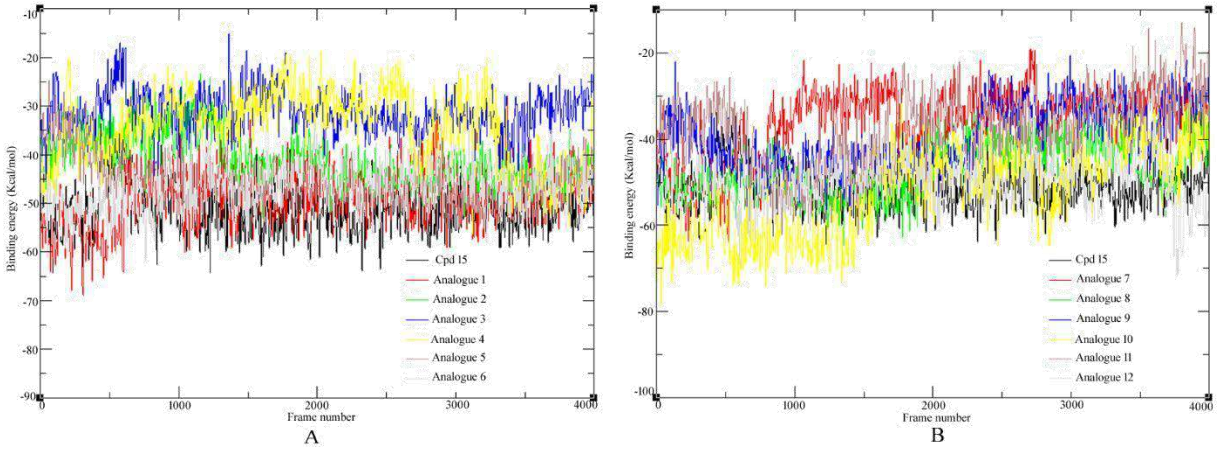


Fig 5:- Binding energy vs frame number of Cpd 15 and *de novo* designed 12 analogues

IV. CONCLUSION

The de novo designing of new and novel chemical entities is a crucial and pivotal approach in the field of drug discovery research. Selected all 34 molecules were successfully docked to cysteine protease falcipain-2 protein with binding affinity range of -4.9 to -9.5 kcal/mol which clearly explained that all inhibitors consist strong binding affinity towards the protein. Best docked complexes (Binding energy < -8 kcal/mol) were furthers used *de novo* design purpose. Initially a number of analogues were developed and best twelve analogues selected based on PLANT dock score and binding interactions. All analogues were found to be shows a number of binding interactions with catalytic amino residues at the active site cavity. High negative PLANT dock score clearly suggested that selected analogues have strong binding affinity towards the protease falcipain-2 protein. The drug ability properties analysis of all analogues using Osiris Property Explorer Tool and SwissADME were explained that selected analogues posses all characteristics to being potential drug molecules. The toxicity assessment also revealed that no indication of mutagenicity or tumerigeniticity were found for the selected analogues. A 40ns of time span MD simulation was performed for analogues-protein complexes. The RMSD and RMSF portrayed undoubtedly explained that all analogues-protease falcipain-2 protein complexes were achieved stability during the MD simulation. The binding energy of most active compound of the dataset and selected analogues were calculated using MM-GBSA approach. Analogue 10 showed higher binding affinity in comparison to Cpd 15, while Analogue 1, Analogue 6, Analogue 8 and Analogue 12 were exhibited comparable binding energy to the Cpd 15. Hence, it can be conclude that Analogue 1, Analogue 6, Analogue 8, Analogue 10 and Analogue 12 have great potential to be anti-malarial agent. For further confirmation, experimental validation such as *in vitro* study will require.

V. ACKNOWLEDGMENT

Author would like to thanks Department of Pharmaceutical Sciences, Pacific Academy of Higher Education and Research University, Udaipur Rajasthan, India-313003 and Maharashtra College of Pharmacy, Nilanga , Maharashtra India-413521 for providing laboratory as will as library requirements during the study. Also thanks Mr. Biologists, Pune for helping out in designing of analogue and docking study.

REFERENCES

[1]. Mital, A., Recent advances in antimalarial compounds and their patents. Curr Med Chem 2007, 14 (7), 759-73.
[2]. World Malarial Report; World Health Organization:. http://www.who.int/malaria/publications/world-malaria-report-2017/en/.
[3]. World Malaria Report 2012.pdf. http://www.who.int/malaria/publications/world_malaria_report_2012/wmr2012_full_rep ort.pdf.
[4]. Fidock, D. A., Drug discovery: Priming the antimalarial pipeline. Nature 2010, 465 (7296), 297-8.
[5]. Nosten, F.; White, N. J., Artemisinin-based combination treatment of falciparum malaria. Am J Trop Med Hyg 2007, 77 (6 Suppl), 181-92.
[6]. Yeung, S.; Pongtavornpinyo, W.; Hastings, I. M.; Mills, A. J.; White, N. J., Antimalarial drug resistance, artemisinin-based combination therapy, and the contribution of modeling to elucidating policy choices. Am J Trop Med Hyg 2004, 71 (2 Suppl), 179-86.
[7]. Sinclair, D.; Zani, B.; Donegan, S.; Olliaro, P.; Garner, P., Artemisinin-based combination therapy for treating uncomplicated malaria. Cochrane Database Syst Rev 2009, (3), CD007483.
[8]. Shah, F.; Mukherjee, P.; Gut, J.; Legac, J.; Rosenthal, P. J.; Tekwani, B. L.; Avery, M. A., Identification of novel malarial cysteine protease inhibitors using structure-based virtual screening of a focused cysteine protease inhibitor library. J Chem Inf Model 2011, 51 (4), 852-64.
[9]. Sajid, M.; McKerrow, J. H., Cysteine proteases of parasitic organisms. Mol Biochem Parasitol 2002, 120 (1), 1-21.
[10]. Leung-Toung, R.; Zhao, Y.; Li, W.; Tam, T. F.; Karimian, K.; Spino, M., Thiol proteases: inhibitors and potential therapeutic targets. Curr Med Chem 2006, 13 (5), 547-81.
[11]. Pandey, K. C.; Dixit, R., Structure-function of falcipains: malarial cysteine proteases. J Trop Med 2012, 2012, 345195.
[12]. Barrett, A. J.; Rawlings, N. D., Evolutionary lines of cysteine peptidases. Biol Chem 2001, 382 (5), 727-33.
[13]. Wu, Y.; Wang, X.; Liu, X.; Wang, Y., Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite. Genome Res 2003, 13 (4), 601-16.
[14]. Ang, K. K.; Ratnam, J.; Gut, J.; Legac, J.; Hansell, E.; Mackey, Z. B.; Skrzypczynska, K. M.; Debnath, A.; Engel, J. C.; Rosenthal, P. J.; McKerrow, J. H.; Arkin, M. R.; Renslo, A. R., Mining a cathepsin inhibitor library for new antiparasitic drug leads. PLoS Negl Trop Dis 2011, 5 (5), e1023.
[15]. Desai, P. V.; Patny, A.; Gut, J.; Rosenthal, P. J.; Tekwani, B.; Srivastava, A.; Avery, M., Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory. Journal of medicinal chemistry 2006, 49 (5), 1576-84.
[16]. Desai, P. V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M., Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The ChemBridge database. Journal of medicinal chemistry 2004, 47 (26), 6609-15.
[17]. Micale, N.; Kozikowski, A. P.; Ettari, R.; Grasso, S.; Zappala, M.; Jeong, J. J.; Kumar, A.; Hanspal, M.; Chishti, A. H., Novel peptidomimetic cysteine protease inhibitors as potential antimalarial agents. Journal of medicinal chemistry 2006, 49 (11), 3064-7.
[18]. Verissimo, E.; Berry, N.; Gibbons, P.; Cristiano, M. L.; Rosenthal, P. J.; Gut, J.; Ward, S. A.; O'Neill, P. M., Design and synthesis of novel 2-pyridone peptidomimetic falcipain 2/3 inhibitors. Bioorganic & medicinal chemistry letters 2008, 18 (14), 4210-4.
[19]. Kerr, I. D.; Lee, J. H.; Pandey, K. C.; Harrison, A.; Sajid, M.; Rosenthal, P. J.; Brinen, L. S., Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: implications for substrate specificity. Journal of medicinal chemistry 2009, 52 (3), 852-7.
[20]. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry 2010, 31 (2), 455-61.
[21]. Douguet, D., e-LEA3D: a computational-aided drug design web server. Nucleic acids research 2010, 38 (Web Server issue), W615-21.
[22]. Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S., LEA3D: a computer-aided ligand design for structure-based drug design. Journal of medicinal chemistry 2005, 48 (7), 2457-68.
[23]. Korb, O.; Stutzle, T.; Exner, T. E., Empirical scoring functions for advanced protein-ligand docking with PLANTS. Journal of chemical information and modeling 2009, 49 (1), 84-96.
[24]. Daina, A.; Michielin, O.; Zoete, V., SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.
[25]. Scientific reports 2017, 7, 42717.
[26]. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A., Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 2001, 30, 211-43.
[27]. Mark, P.; Nilsson, L., Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. The Journal of Physical Chemistry A 2001, 105 (43), 9954-9960.
[28]. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.;

Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). J Comput Chem 2005, 26 (16), 1752-80.

[29]. Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C., A Limited Memory Algorithm for Bound Constrained Optimization. SIAM Journal on Scientific Computing 1995, 16 (5), 1190– 1208.

[30]. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000, 33 (12), 889-97.

[31]. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 2011, 51 (1), 69-82.

[32]. Durham, E.; Dorr, B.; Woetzel, N.; Staritzbichler, R.; Meiler, J., Solvent accessible surface area approximations for rapid and accurate protein structure prediction. J Mol Model 2009, 15 (9), 1093-108.