Molecular Docking and QSAR Study of Chalcone and Pyrimidine Derivatives as Potent Anti-Malarial Agents against *Plasmodium falciparum*

Dayena J. Christian^{1,a}, Rajesh H. Vekariya^{1,b}, Kinjal D. Patel^{1,c},
Dhanji P. Rajani^{2,d}, Smita D. Rajani^{2,e}, Kailash Parmar^{3,f},
Mahesh T. Chhabaria^{2,g*}, Hitesh D. Patel^{1,h*}

¹Department of Chemistry, School of Sciences, Gujarat University, Navrangpura, Ahmedabad 380009, Gujarat, India.

²Microcare Laboratory and TRC, Surat, Gujarat, India

³Department of Pharmaceutical Chemistry, L. M. College of Pharmacy, Navrangpura, Ahmedabad 380009, Gujarat, India

^adyna.denish@gmail.com, ^brajeshvekariya007@gmail.com, ^ckinjalpatel.chem@gmail.com,
^dmicrocaresurat@yahoo.co.in, ^emicrocaresurat@yahoo.com, ^fkailashparmar@ymail.com,
^gmtchhabaria@yahoo.com, ^hdrhiteshpatel1@gmail.com

Keywords: Molecular Docking, QSAR, Antimalarial activity, *Plasmodium falciparum*, pyrimidine derivatives, chalcone derivatives

Abstract. A data set of chalcone and pyrimidine derivatives with anti-malarial activity against *Plasmodium falciparum* was employed in investigating the quantitative structure-activity relationship (QSAR). Molecular docking study was performed for *Plasmodium falciparum* dihydrofolate reductase (*PfDHFR-TS*). Genetic function approximation (GFA) technique was used to identify the descriptors that have influence on anti-malarial activity. The most influencing molecular descriptors identified include thermodynamics, structural and physical descriptors. Generated model was found to be good based on correlation coefficient, LOF, rm^2 and rcv^2 values. Nrotb, solubility, polarizability may have negative influence on antimalarial activity or play an important role in growth inhibition of *Plasmodium falciparum*. The QSAR models so constructed provide fruitful insights for the future development of anti-malarial agents.

1. Introduction

Malaria is the most common parasitic disease in the tropic and sub-tropic regions of the world today. The World Health Organization's 2013 'World Malaria Report' estimates that The disease still took an estimated 6,27,000 lives in 2012. From total the reported cases of malaria in the year of 2012 in south-east Asia region, 52% cases were only from India [1]. Nearly all antimalarial drugs were developed because of their action against asexual erythrocytic forms of malaria parasites that cause clinical illness. Malaria is caused by *Plasmodium* parasites like *P. falciparum*, *P. vivax*, *P. malariae*, and *P. oval* [2], of which the most lethal is *Plasmodium falciparum* (*Pf*). The biology of this parasite is well understood with the help of genomic [3] and proteomic projects [4]. Some important antimalarial compounds are as follows:

Dihydrofolatereductase (DHFR) is responsible for the reduction of 5,6 double bond of dihydrofolate to tetra hydrofolate through a catalytic cycle in malaria parasite. Models of *p**f*DHFR have also been reported based on the homology of the DHFR sequences [5, 6, 7]. In *Plasmodia*, DHFR and TS coexist as a single-chain bifunctional enzyme, in contrast to prokaryotes and higher eukaryotes where the two proteins are distinct monofunctional enzymes. In bifunctional DHFR-TS, individual DHFR and TS domains have polypeptide folds closely related structurally to those of their respective monofunctional counterparts [8]. Before the release of *Pf*DHFR-TS crystal structures, researchers developed homology models to study the resistance mechanism of commonly used antifolate drugs [7]. Drug resistance acquired by the malarial parasites is a major problem in the treatment and control of malaria. Dihydrofolatereductase (DHFR) in *Pf* (*Pf*DHFR-

TS; TS refers to thymidylate synthase bound to DHFR in *Pf*), which catalyzes the reduction of dihydrofolate to tetrahydrofolate, is one of the most widely studied enzymes in antimalarial drug design due to its potential role in DNA synthesis [9, 10].

Christian et al. was synthesized a series (3, 5 and 7) in the strong basic condition via under microwave irradiation, were give good results with *plasmodium falciparum* (*Pf*). They have synthesized prop-2-en-1-ones (3a-3g) by using heterocyclic/aromatic ketone and aryl/cyclic aldehydes in 10-15 min which further react with guanidine hydrochloride to yield pyrimidin-2-amines (5a-5g) in 25-30 min. Series 3 was also being a precursor for the pyrimidin-2-thiols (7a-7g) by utilizing thiourea in 35-40 min (**Scheme 1**). All series compounds tested for *pf*, from which 5a, 5b, 5c, 5f, 5e and 7c shows very good results when compared with standard [11].

Scheme 1: Reaction scheme.

2. Materials and Method

Data set

The ligand set of twenty one compounds were derived as previously reported method by Christian et al. in 2014 [11], were used for MolDock study. Inhibitory concentration of all molecules was expressed as IC₅₀. The general chemical structures and biological activity values of all of the compounds are shown in **Table 1**.

Table 1: Energetic distribution and IC₅₀ values of substituted pyrimidines.

Ligand [11]	Scaffold [11]	IC ₅₀ (μg/mL) [11]	MolDock Score (kcal/mol)	Ligand [11]	Scaffold [11]	IC ₅₀ (μg/mL) [11]	MolDock Score (kcal/mol)
3a		0.56	-88.942	5e		0.23	-89.417
3b		0.93	-89.307	5f		0.26	-96.140
3c		0.64	-88.367	5g		0.98	-90.485
3d		0.58	-85.763	7a		0.84	-85.189
3e		1.33	-89.211	7b		1.45	-94.220
3f		1.12	-92.350	7c		0.078	-99.110
3g		0.94	-80.868	7d		0.96	-93.302
5a		0.26	-89.393	7e		0.84	-87.554
5b		0.124	-93.619	7g		0.32	-88.022
5c		0.028	-97.302				
5d		0.54	-87.865				

3. Molecular Docking

3.1. Binding site prediction

It is well known that plasmodium falciparum dihydrofolatereductase (*PfDHFR-TS*) possess a catalytic site for enzymatic activity [12]. Molegro virtual docker (MVD) ‘Cavity’ module was used to recognize ligand binding potential sites using an interaction energy scheme to find energetically favorable binding sites between the protein and a van der Waals probe [13]. The cavities clustered by spatial proximal scheme were utilized to rank the top scoring predicted cavities. The experimentally known active site of *PfDHFR-TS* (Phe96) [12] was deployed as docking cavity.

3.2. Molecular docking of substituted prop-2-en-1-ones and pyrimidines and their energetic profile

Prior to docking, the ligand dataset was energy minimized using the Gchemical all-atom force field engineered in the PyRx program [14]. The crystal structure of *PfDHFR-TS* (entry: 4DPD) was downloaded from Protein Databank (PDB). The protein structures were preprocessed by removing crystallographic waters and ions. The docking simulations were performed using Molegro Virtual Docker (MVD) 8.0 program. MVD develops ligand poses according to the search space in the protein binding cavity by utilizing genetic evolutionary methods (GA). Docking was carried out with the following customized settings: Grid resolution = 0.30, population size = 50, number of generations = 10 and the number of solutions = 10. The best poses for each ligand was selected by initially sorting out the MolDock score and then by H bond energy term.

3.3. Model validation and stereochemistry corrections (530 AA)

Ramachandran plot [15] exhibited 95.7% amino acids in core areas of ϕ - ψ distributions, with 4.0% dispensed in additional allowed regions and 0.4% outliers (**Figure 1A**). Besides, the PROSA-WEB Z-score (a composite scoring function consisting of the following pseudo energy terms: C β interaction, all-atom pairwise, solvation, torsional angle and agreement measures: secondary structure and solvent accessibility) [16] of the modeled *PfDHFR-TS* protein was observed to be -9.77, which is quite close to that of the experimental X-ray structures (**Figure 1B**). Thus, the model was proved to be structurally realistic and was thus used for further analysis.

Figure 1: A. Model quality (a black dot on the Z-score plot) resembles experimental structure statistics. B. Ramchandran plot showing amino acid distribution in a reliable structure.

3.4. Interaction of pyrimidine derivatives in *PfDHFR-TS*.

A cavity with active site residues, along with adjacent neighbors within a proximity of 6 Å, was constructed to implement as a docking grid. A grid space with a site volume of 209 Å³ predicted by MVD [17] was utilized in docking simulations. The ligand dataset consisted of twenty-one pyrimidine derivatives. We tried to investigate the binding confirmation of the substituted pyrimidine moiety initially to distinguish the floor and to allow the terminal groups to interact with the active site residues (**Figure 2**). It is evident that the major pharmacological activity of pyrimidine is due to its lipophilicity (predicted LogP=-0.312) [18].

The MolDock score of the ligand dataset constituted major contribution of van der Waals contacts with few H bonds. The substituted pyrimidine moiety is enriched with H-bond acceptors, H-bond donors and hydrophobic pharmacophore groups, and the active site of *PfDHFR-TS* possesses chargeable and non-polar amino acids, there is a possibility to develop H bonding and other modes of interactions (**Table 2**). The substituted pyrimidine moiety interacts with the active site residues.

Table 2: Interaction between ligand proteins by amino acids as given below.

Compound Code	IC ₅₀ value (µg/mL)	Amino acids form Hydrogen bonding	Amino acids form π contact
5a	0.26	Asp12	Lys321, Tyr326, Tyr322, His323, Phe293, Gln327, Asn330
5b	0.124	Asp12	Lys321, Tyr326, Tyr322, His323, Phe293, Gln327, Asn330
5c	0.028	Tyr365, Lys359, Gln327	Tyr326, Tyr322, Pro324, His323, Asp212
5e	0.23	Tyr365, Lys359, Gln327	Tyr326, Tyr322, Pro324, His323, Asp212
5f	0.26	Tyr322, His323	Asp212, Pro324, Tyr365, Lys359, Gln542
7e	0.078	Tyr365, Lys359, Gln327	Tyr322, Asp212, His323, Tyr326, Lys321

A

Figure 2: A. A docked poses of all compounds, B. H bond of 5a with active site amino acids, C. H bond of 5b with active site amino acids, D. H bond of 5c with active site amino acids, E. H bond of 5e with active site amino acids. H bonds are shown in green color dotted lines, F. H bond of 5f with active site amino acids. H bonds are shown in green color dotted lines, G. H bond of 7e with active site amino acids. H bonds are shown in green color dotted lines.

4. Qsar Study

4.1. Data set

Potent antimalarial activity has been reported in a series of chalcones and pyrimidine by Christian [11] against *Pf* (**Tables 1**). This data was used for QSAR study. The IC₅₀ values were converted to the negative logarithmic scale (pIC₅₀) to achieve normal distribution shown in **Table 3**. The strategy for the selection of compounds to be included in the test set was a random selection of compounds that exhibited a varied range of inhibitory activities and structural diversities. Twenty compounds were randomly divided into training set and test set of fifteen and five compounds, respectively.

Table 3: Biological activity of training set and test set compounds.

Compound no.	pIC50	Compound no.	pIC50	Compound no.	pIC50
3a	8.585	5a	8.991	7a	8.51
3b	8.375	5b	9.319	7b	8.281
3c	8.504	5c	9.939	7c	9.524
3d	8.569	5d	8.674	7d	8.451
3e	8.132	5e	8.979	7e	8.45
3f	8.347	5f	9.044	7f	8.371
3g	8.347	5g	8.403	7g	8.921

4.2. Computation and selection of molecular descriptors

Descriptors are considered as vital element of any ligand based study, due to any biological property can vary depends on molecular property of chemical structures. In this study, molecular descriptors representing structural, thermodynamic and other properties were calculated using Microsoft excel 2010. To develop QSAR models descriptors were selected by analysis of correlation matrix. From correlation matrix, highly correlated descriptors with value above 0.75 (implying highest multicollinearity) and descriptors with value zero were removed from the study. Descriptors used in both studies are listed in the **Table 4**. Calculated descriptors value (for Eq. 1) used in the QSAR study of *P. falciparum* are given in **Table 5**. Correlation matrices of descriptors (for Eq. 1) used in the QSAR study of *P. falciparum* are given in the **Table 6**.

Table 4: List of the descriptors used in the study.

Structural	Molecular weight, number of hydrogen bond donors, number of hydrogen bond acceptors, Rigid double bond, Number of aromatic ring, Number of hydrogen bond
Thermodynamic	Log of partition coefficient, molar refractivity, Surface tension, Volume, Topological polar surface area, Solubility, Molar volume,
Others	Number of rotational bonds, Number of atoms, Perachor, Polarizability, Aqueous solubility of compound

Table 5: Descriptors value of training set and test set compounds for *P. falciparum* (used in the equation 1).

Compound no.	Nrotb	solubility	Polarisability
3a	3	-3.17	26.35
3b	3	-3.95	26.46
3c	3	-3.62	24.05
3e	4	-3.16	21.11
3f	3	-4.67	29.04
3d	3	-3.14	26.35
5a	2	-3.42	29.81
5b	2	-4.45	29.93
5c	2	-3.88	27.52
5d	2	-3.73	29.81
5g	3	-4.27	27.88
7a	2	-3.95	31.37
7c	2	-4.41	29.08
7e	3	-3.98	26.86
7d	2	-4.26	31.37

Table 6: Correlation matrix of descriptors used in Eq. 1 for *P. falciparum*.

	Nrotb	Solubility	Polarizability
Nrotb	1		
Solubility	0.392302	1	
Polarizability	-0.85696	-0.5746	1

4.3. Validation test

Statistical significance of relationship between chemical structural descriptors and antimalarial activity was analysed by variance inflation factor (VIF) analysis, cross validation, or internal validation with training set and external validation with external data set.

To check the inter-correlation of descriptors variance inflation factor (VIF) analysis was performed. VIF value is calculated from $1/(1 - r^2)$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. VIF value is larger than 10 indicates that information of descriptors can be hidden by correlation of descriptors and multicollinearity [19, 20].

4.4. Evaluation of predictive ability of QSAR models

Evaluation and predictive power of the models were estimated using external dataset by examination of the statistical parameters such as r_{pred}^2 , r_m^2 .

Figure 3: Graph of observed and predicted activities of the training set compounds by Eq. 1 (*P. falciparum*).

Figure 4: Graph of observed and predicted activities of the test set compounds by Eq. 1 (*P. falciparum*).

Predictive correlation coefficient values (r_{pred}^2) is the predictive ability of each analysis was determined from a set of compounds that were not included in the training set. The predictive r^2 (r_{pred}^2) value is based on molecules of the test set only and is defined as [21]:

$$r_{pred}^2 = (SD - PRESS)/SD$$

Where SD is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules and PRESS is sum of the squared deviation between predicted and actual activity values for every molecule in the test set.

To better understand the external predictability of models, modified r^2 (r_m^2) was determined by the following equation [22]:

$$r_m^2 = r^2 [1 - \frac{r^2}{n-2}]$$

Where r^2 is the squared correlation coefficient between observed and predicted values and r_0^2 is the squared correlation coefficient between observed and predicted values without intercept.

5. Results and Discussions

Table 7: List of QSAR equations.

Equation no.	Equation
1	$PIC50 = 17.133 + (-1.2799 * Nrotb) + (-0.299 * solubility) + (-0.225 * Polarizibility)$
2	$PIC50 = 10.749 + (-0.5479 * Nrotb) + (0.2479 * solubility) + (0.123 * C \text{ Logp})$
3	$PIC50 = 13.247 + (-0.873 * Nrotb) + (-0.0189 * MR) + (-0.0297 * Polarizibility)$

Table 8: Summary of statistical parameters.

	r^2	r_0^2	rm^2	VIF
3g	0.990292	0.994251	0.927975	103.0053
5e	0.995514	0.994251	0.991976	222.9075
5f	0.99839	0.994251	0.934164	621.2885
7b	0.979857	0.994251	0.862315	49.64485
7g	0.973964	0.994251	0.835272	38.40866

Table 9: Observed activity and predicted activity of training set compounds as per Eq. 1.

Compound no.	Predicted pIC50	Actual pIC50	Residual
3a	8.322041809	8.585	0.262958
3b	8.530912492	8.375	-0.15591
3c	8.973574417	8.504	-0.46957
3d	8.313057738	8.569	0.255942
3e	8.216447311	8.132	-0.08445
3f	8.166847331	8.347	0.180153
5a	8.899445113	8.991	0.091555
5b	9.18093622	9.319	0.138064
5c	9.551725578	9.939	0.387274
5d	8.992280512	8.674	-0.31828
5g	8.307692324	8.403	0.095308
7a	8.707657782	8.51	-0.19766
7c	9.359938247	9.524	0.164062
7d	8.800493181	8.451	-0.34949
7e	8.450023251	8.45	-2.3E-05

Table 10: Observed activity and predicted activity of test set compounds as per Eq. 1.

Compound no.	Predicted pIC50	Actual pIC50	Residual
3g	7.855376841	8.347	0.491623
5e	8.644805272	8.979	0.334195
5f	8.843823272	9.044	0.200177
7b	8.98914889	8.281	-0.70815
7g	8.115904993	8.921	0.805095

To gather information about which physicochemical properties of compounds influencing the antimalarial activity, 100 QSAR equations were generated with 23 preselected descriptors from correlation matrix using GFA technique. Results of the best models with four parameters (**Table 7**) as per the rule of “per descriptors five compounds,” which showed acceptable statistical characteristics, for *P. falciparum* (**Table 8**) (**Figure 3**, **Figure 4**).

For any QSAR study, it is necessary that molecular features affecting the biological activity should not be inter-correlated with each other. VIF analysis of both studies indicates low inter-correlation of the descriptors used in the selected models. For example, correlation matrix value of all the descriptors are calculated in Table 5. Model 1 is considered to be the best based on internal and external validation. The regression correlation coefficient of 0.90 shows good correlation with biological activity, cross-validation coefficient of 0.67 indicates good internal predictability, predicted correlation coefficient (r^2_{pred}) of 0.67 and r^2_{m} value of 0.90 indicates reliability and significance of the model as external validity parameters. Each equation is assessed by Friedman's lack of fit. Molecular descriptors used in the model 1 are Nrotb, solubility, polarizability (**Table 9**). In which, polarizability to be important between the electrons in a ligand and a biological receptor. Likewise solubility and Nrotb of the compounds are also played an important role in the changes in biological activity (**Table 10**).

6. Conclusion

Molecular docking study shows that the substituted pyrimidine moiety have good binding in *PfDHFR-TS* pocket and displayed excellent antimalarial activity as compared to chalcone derivatives. The QSAR study shows that structural, thermodynamic and physical descriptors are responsible for describing activity of chalcones and 2-substituted pyrimidine derivatives. The QSAR model is statistically and chemically sound with excellent predictive power as evidenced from the predicted activity of test set compounds. It shows that as the Nrotb, solubility, polarizability may have negative influence on antimalarial activity. The results showed that the 2-substituted pyrimidine derivatives having lower values of affected descriptors, which demonstrated their excellent antimalarial activity as compared to chalcone derivatives.

7. Acknowledgement

We are grateful to L. M. college of Pharmacy for providing technical and software facilities for the QSAR study. The authors are thankful to the Department of Chemistry, Gujarat University, Ahmedabad, for providing the necessary facilities. UGC-Info net & INFLIBNET Gujarat University are acknowledged for providing the e-source facilities. R.H.V. is thankful to UGC-BSR (F.7-74/2007 (BSR)) for financial assistance.

References

- [1] World Health Organisation, (2013) World Malaria Report.
- [2] P.K. Chiang et al., Malaria: therapy, genes and vaccines, Curr. Mol. Med. 6 (2006) 309-326.
- [3] M.J. Gardner et al., Genome sequence of the human malaria parasite *Plasmodium falciparum*, Nature 419 (2002) 498-511.
- [4] L. Florens et al., A proteomic view of the *Plasmodium falciparum* life cycle, Nature 419 (2002) 520-526.
- [5] T. Lemcke, I.T. Christensen, F.S. Jorgensen, Towards an understanding of drug resistance in malaria: three-dimensional structure of *Plasmodium falciparum* dihydrofolate reductase by homology building, Bioorg. med. chem. 7 (1999) 1003-1011.

-
- [6] G. Rastelli et al., Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with *Plasmodium falciparum* dihydrofolate reductase: structural basis of antifolate resistance, *Bioorg. med. chem.* 8 (2000) 1117-1128.
- [7] R.T. Delfino, O.A. Santos, J.D. Figueroa-Villar, Molecular modeling of wild-type and antifolate resistant mutant *Plasmodium falciparum* DHFR, *Biophys. Chem.* 98 (2002) 287-300.
- [8] D.R. Knighton et al., Structure of and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase, *Nat. Struct. Biol.* 1 (1994) 186-194.
- [9] A. Gregson, C.V. Plowe, Mechanisms of resistance of malaria parasites to antifolates, *Pharmacol. Rev.* 57 (2005) 117-145.
- [10] I.M. Kompis, K. Islam, R.L. Then, DNA and RNA synthesis: antifolates, *Chem. Rev.* 105 (2005) 593-620.
- [11] D.J. Christian et al., Microwave Assisted Synthesis and in Vitro Anti-malarial Screening of Novel Pyrimidine Derivatives, *World J. Pharma. Pharma Sci.* 3 (2014) 1955-1971.
- [12] J.M. Beierlein, N.G. Karri, A.C. Anderson, Targeted Mutations of *Bacillus anthracis* Dihydrofolate Reductase Condense Complex Structure– Activity Relationships, *J. Med. Chem.* 53 (2010) 7327-7336.
- [13] A.T.R. Laurie, R.M. Jackson, Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites, *Bioinformatic* 21 (2005) 1908-1916.
- [14] L.K. Wolf, Quidditch For Chemists, *ChemEng News Arch* 87 (2009) 48-48.
- [15] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Conformation of polypeptides and proteins, *J. Mol. Biol.* 7 (1963) 95-99.
- [16] P. Benkert, S.C.E. Tosatto, D. Schomburg, QMEAN: A comprehensive scoring function for model quality assessment, *Proteins Struct. Funct. Bioinf.* 71 (2008) 261-277.
- [17] S.F. Altschul et al., Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403-410.
- [18] <http://www.molinspiration.com>.
- [19] I.A. Khan et al., Quantitative structure–activity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors, *Eur. J. Med. Chem.* 44 (2009) 229-238.
- [20] S. Peterangelo, P. Seybold, Synergistic interactions among QSAR descriptors, *Int. J. Quantum Chem.* 96 (2004) 1-9.
- [21] S. Kulkarni, V.M. Kulkarni, Three-Dimensional Quantitative Structure-Activity Relationship of Interleukin 1- β Converting Enzyme Inhibitors: A Comparative Molecular Field Analysis Study, *J. Med. Chem.* 42 (1999) 373-380.
- [22] P. Roy, K. Roy, On some aspects of variable selection for partial least squares regression models, *QSAR Comb. Sci.* 27 (2007) 302-313.