Theoretical studies of Thiazolyl-Pyrazoline derivatives as promising drugs against malaria by QSAR modelling combined with molecular docking and molecular dynamics simulation

Arwansyah Arwansyah, Abdur Rahman Arif, Gita Syahputra, Sukarti Sukarti & Isman Kurniawan

Published online: 07 Jun 2021.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Check for updates

Theoretical studies of Thiazolyl-Pyrazoline derivatives as promising drugs against malaria by QSAR modelling combined with molecular docking and molecular dynamics simulation

Arwansyah Arwansyah [a], Abdur Rahman Arif [b], Gita Syahputra [c], Sukarti Sukarti[a] and Isman Kurniawan [d,e]

[a]Department of Chemistry, Faculty of Science, Universitas Cokroaminoto Palopo, Palopo, Indonesia; [b]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, Indonesia; [c]Research Center for Biotechnology, Indonesian Institute of Science, Bogor, Indonesia; [d]School of Computing, Telkom University, Bandung, Indonesia; [e]Research Center of Human Centric Engineering, Telkom University, Bandung, Indonesia

ABSTRACT

We investigate thiazolyl-pyrazoline derivatives as promising drugs for the anti-malarial. Protein kinase G is a primary target for treating malaria due to its essential role in *Plasmodium falciparum* life cycle. In this present study, several computational approaches such as QSAR modelling, molecular docking, and all-atom MD simulation are performed to screen 36 drug candidates against malaria. From QSAR analysis, three potent drugs are selected based on the strong correlation between the inhibitory action, i.e. pEC50 and various descriptors. Further, those selected drugs are used as ligand molecules for molecular docking. We predict three complexes of models 1, 2, and 3 bind to the catalytic site of protein kinase G, suggesting those ligands may become potent inhibitors for *Plasmodium falciparum*. To validate the structural stability of those complexes, the parameters of RMSD, RMSF, and Rg are calculated from MD simulation. All models are stable along simulation since no significant fluctuations are observed in those validity parameters. Moreover, the binding energy is estimated at each model using MM-GBSA method and model 2 becomes the most stable structure. Finally, three ligands are assumed to have potential as inhibitors and the ligand of model 2 may become the most promising drug against malaria.

1. Introduction

Malaria transmitted by a bite of infected mosquitoes is one of the serious global diseases that has caused 400,000 deaths and hundreds of million people of new infection each year [1]. Therefore, searching for promising anti-malarial drugs for the treatment of the disease can be a crucial challenge due to the increasing cases affected by *Plasmodium* parasites. In the molecular investigation, inhibition of the parasite life is required for preventing a new infection to the human. A lot of proteins are directly connected to the *Plasmodium* life [2–6]. Protein kinase is one of the Plasmodium proteins that can be considered an important target because of the critical role in the parasite life cycle. This protein relates to a complex system of second messenger signalling, phosphoinositide metabolism, calcium transport, etc. [2, 3, 7]. It indicates that protein kinase is implicated in calcium mobilisation, parasite signalling, and even transporting the parasite invasion to red blood cells.

In experimental research, the inhibitory activities of several anti-malarial drugs to protein kinase have been investigated. In the paper presented by Rama and co-workers, a series of pyrazolopyrimidine and imidazopyrazine can become anti-malaria drugs by in vitro analysis due to the inhibitory action to *Plasmodium falciparum* calcium-dependent protein kinase 4 (*PfCDPK4*) [3]. Also, protein kinases as targets for anti-malarial intervention have been reported by Doerig and co-workers [7].

The authors suggest that the inhibition of protein kinases is needed for controlling a new infection by the parasites. From viewpoints of theoretical investigation, the inhibitory activities by the interaction of drug candidates in complex with protein kinase have been presented. For example, plasmepsin-II inhibitor di-tertiary amines by molecular docking combined with QSAR method have capabilities as anti-malarial agents since those drug candidates bind to the enzyme's active site [8]. Potential drugs against malaria such as imidazopyrazine and its analogs by molecular docking and MD simulations have reported the inhibition activities to phosphatidylinositol-4-OH kinase type III beta (*pfPI4KB*). The test compounds of ZINC78988474 and ZINC20564116 are identified as potent *pfPI4KB* inhibitors due to the lowest binding energies from all candidates [9]. A paper presented by Marilia and co-workers [10] has also investigated quinazoline derivatives as potential drugs as anti-malaria. The authors employ machine learning techniques combined with molecular docking to screen a series of compounds. Thus, the inhibition of a protein kinase of *Plasmodium* parasites becomes a crucial target for designing a new drug to treat malaria.

Cyclic guanosine-3,5-monophosphate-dependent protein kinase or protein kinase G, one of the kinds of protein kinase of *Plasmodium falciparum*, is selected as a primary target for anti-malaria drugs. Bakkouri and co-workers have investigated

the tertiary structure of protein kinase G by X-ray analysis [2]. This protein consists of various domains such as cyclic nucleotide-binding (CNB-A, B, C, D), kinase domain C-lobe, kinase domain N-lobe and auto-inhibitory segment (AIS). The activation of this protein involves the binding of cyclic guanosine-3,5-monophosphate (cGMP) to a segment of protein kinase G. This holoenzyme bounds to several amino acid residues of protein kinase G, including VAL105, LYS113, MET115, PHE121, GLY122, GLU123, ALA124, ARG132, SER133, and ILE136. Hence, favourable drug candidates that can attach to cGMP sites are assumed to have an inhibitory action to protein kinase G.

In this research, we perform quantitative structure-activity relationship (QSAR) modelling to screen a series of thiazolyl-pyrazoline derivatives as promising drugs against malaria. This technique is one of the favourable and innovative methods in the field of drug design and molecular modelling by optimising and validating the relationship between a compound and its chemical properties [11–18]. To find the binding site, including the orientation pose of anti-malaria drugs with a receptor, we perform the molecular docking simulation on the thiazolyl-pyrazoline/protein kinase G complex. This method is also widely used as a drug discovery because it can explain a particular small molecule in binding with a target protein by using a searching algorithm and scoring functions [19–21]. To confirm the complex obtained by molecular docking is stable in the water solvent, all-atom molecular dynamics (MD) simulation is performed on thiazolyl-pyrazoline in complex with protein kinase G. Moreover, the binding energy of the complex is estimated from the trajectories of MD simulation. A combination of some computational approaches, i.e. QSAR, molecular docking, and MD, is expected not only to present a better understanding of thiazolyl-pyrazoline/protein kinase G interactions but also to occupy notable significance for further design of new potent drugs for the treatment of malaria.

2. Materials and methods

2.1. QSAR modelling

2.1.1. Data preparation and selection
To perform QSAR modelling on thiazolyl-pyrazoline derivatives, we retrieve the dataset, including the value of inhibitory potency (EC50) of the compounds for 36 structures from Ref. [22]. The chemical structures of those compounds are provided in an additional file. 36 structures are created by Avogadro program packages. Next, those structures are optimised by using semi-empirical calculations (AM1) in the gas phase to obtain the stable geometry of each molecule. Then, all structures are saved in Mol2 format. Meanwhile, inhibitory potency

(EC50) of 36 structures is converted to negative logarithm values $\log_{10}(1/EC_{50})$ or written to pEC50. This data is used to validate the QSAR analysis in further analysis.

To obtain the descriptor in relation to the compound's structural, topological, and chemical properties, Padel program package is employed on thiazolyl-pyrazoline derivatives [23]. Initially, a total of 1440 descriptors are obtained after calculations. Since large descriptors are collected, we need to decide the descriptors with strong correlations with inhibitory activity pEC50 (mM). The number of descriptors is reduced by applying some step selections. In the first selection, the descriptor with constant or almost constant values is discarded. The descriptor which contains *NaN* values and empty features is erased. Also, the descriptor with a strong correlation with other descriptors (Pearson correlation coefficient >0.8) is removed. From this procedure, the dataset of the generated descriptor remains 72 features. Then, we apply the Genetic Algorithm (GA) method to select the most important descriptors from the primary variable collection. The GA protocols are set according to the similar procedure provided in Ref. [24]. The size of the population and learning generations are set to 200 and 100,000, respectively. Other parameters are arranged as a default form. Finally, three descriptors with strong correlations with inhibitory activity pEC50 (mM) are found from all selection protocols. The name, including the definition of those descriptors, are listed in Table 1.

2.1.2. QSAR method and validation
For QSAR modelling, multiple linear regression (MLR) is commonly used to investigate the linear relationship between a dependent variable, i.e. PEC50, and independent variables corresponding to various descriptors. The data analysed by MLR are represented as a linear equation. The model development is started by splitting 36 compounds into two sets, i.e. training and test data. The training data consists of 28 compounds. Meanwhile, 8 compounds are selected as the test data. The descriptor values of 3 descriptors with high correlations with pEC50 are shown in Table 2. Further, we apply two other methods, i.e. random forest (RF) and support vector machine (SVM), to analyse the correlation between descriptor and inhibitory activity. RF is a general method that randomly selects the training data to examine a large number of features. CART algorithm developed by Breiman and co-workers is basic theory for RF method where the diversity of trees is improved as the learning ability [25,26]. RF can solve a regression or classification issue in some ways. For example, the variable importance of a data set is measured by fitting a random forest to the data, then the value of the features are recorded and averaged to the training data. The standard deviation is found after normalising the feature score in the data set. In this current letter, the descriptors are estimated by RF to find the correlation of given compounds with chemical structure and properties with their inhibitory values as anti-malaria. Another popular method for QSAR modelling is SVM method developed by Vapnik [27]. This method generates some principles such as developing the statistical approach theory, handling to minimise the error value, and producing the optimal whole response [28]. This technique has been widely applied to determine the classification or regression

Table 1. List of selected descriptors and the physical-chemical definition.

Descriptor	Description class	Descriptor definition
AATS7e	Auto correlation descriptor	Average Broto–Moreau autocorrelation - lag 1 / weighted by Sanderson electronegativities
AlogP	Physico-chemical	Logarithm of the octanol-water partision
ATSC4s	Auto correlation descriptor	Average centred Broto–Moreau autocorrelation - lag 0 / weighted by I-state

Table 2. The values of the descriptors, experimental and predicted pEC50 values by MLR, RF and SVM models for training and test sets.

No.	AATS7e	ALogP	ATSC4s	Exp (pEC50)	MLR	RF	SVM
Training set							
1	8.616.535	24.801	6.463.437	4.596	4.602	4.673	4.695
2	7.992.811	35.141	−4.287.536	4.757	4.852	4.834	4.837
3	7.966.531	36.586	−3.884.633	4.825	4.847	4.835	4.831
4	8.046.704	32.495	−15.015.726	4.961	4.966	4.949	4.863
5	8.209.171	33.056	−10.535.654	4.922	4.864	4.936	4.870
6	7.991.385	27.180	−8.575.302	4.944	4.945	4.941	4.850
7	7.763.543	26.144	−4.763.540	4.948	4.980	4.954	4.874
9	8.043.407	26.707	−6.773.654	4.909	4.912	4.918	4.853
10	8.012.786	28.152	−6.419.912	4.868	4.909	4.908	4.852
11	8.106.203	24.061	−15.874.330	4.974	5.006	4.970	4.974
12	8.154.468	24.622	−10.465.366	4.975	4.930	4.967	4.975
13	8.037.675	18.746	−10.829.863	4.974	5.004	4.970	4.974
14	7.853.266	17.710	−3.468.137	4.976	4.987	4.977	4.976
15	7.937.788	19.264	−6.388.219	4.977	4.984	4.971	4.977
16	8.076.718	25.919	−5.116.022	4.974	4.888	4.954	4.974
17	8.040.209	27.364	−4.785.052	4.948	4.888	4.954	4.948
18	8.151.591	23.273	−13.706.227	4.963	4.973	4.966	4.963
20	8.067.944	17.958	−9.189.209	4.962	4.981	4.961	4.962
22	7.950.789	18.476	−4.716.120	4.958	4.966	4.963	4.958
23	7.819.267	33.088	−2.922.553	4.973	4.903	4.967	4.973
24	7.794.927	34.533	−2.645.213	4.961	4.899	4.936	4.961
25	7.869.182	30.442	−10.096.634	4.949	4.980	4.953	4.949
29	7.735.314	25.645	−2.602.887	4.988	4.969	4.974	4.988
30	8.025.044	47.974	−3.016.840	4.623	4.754	4.706	4.623
32	8.074.959	45.328	−10.195.216	4.966	4.831	4.867	4.966
34	8.021.192	40.013	−6.451.876	4.705	4.838	4.798	4.705
35	7.867.306	38.977	2.313.220	4.788	4.797	4.804	4.788
36	7.941.091	40.531	−2.696.843	4.928	4.820	4.817	4.928
Test set							
8	7.902.167	27.698	−3.883.451	4.948	4.919	4.926	4.948
19	8.198.887	23.834	−8.433.738	4.912	4.898	4.960	4.912
21	7.851.358	16.922	−1.965.586	4.958	4.976	4.977	4.958
26	7.940.469	31.003	−4.978.976	4.970	4.900	4.945	4.970
27	7.823.849	25.127	−6.354.929	4.967	4.985	4.972	4.967
28	7.727.330	24.091	2.416.255	4.856	4.926	4.827	4.856
31	8.000.704	49.419	−2.739.318	4.709	4.750	4.716	4.709
33	8.118.809	45.889	−5.077.883	4.706	4.759	4.847	4.706

of the data due to many attractive features and encouraging practical execution. In SVM, several parameters are related to the calculation performance, such as the capacity C parameter, the kernel function, and epsilon-intensitive loss function. Also, the Gaussian radial basis function (RBF) is generally utilised in SVM. Therefore, the usage of those parameters is a necessary step for SVM estimation [29]. In this present letter, the regression assessment of MLR, RF, and SVM is analysed by Python 2.7 packages [30].

To validate the QSAR model, the internal and external validations are performed by utilising the model to predict training and test data. Here, we calculate several parameters to evaluate the validity of the model. Firstly, the coefficients of determination (R^2) are calculated to evaluate the strength of the relationship between the actual values and predicted values. The model performance is also evaluated by considering the slope value (k) and $\frac{(r^2 - r_0^2)}{r^2}$. The r^2 and r_0^2 parameters represent the correlation coefficient between the actual and predicted values with and without intercept, respectively. Finally, we calculate the modified r^2 (r_m^2) to present a better predictive ability of the model [31,32]. The calculation of the validation parameters is formulated in the following equations:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \quad (1)$$

$$k = \frac{\sum (y \times \hat{y})}{\sum (\hat{y})^2} \quad (2)$$

$$r^2 = \frac{\left[\sum (y - \bar{y})(\hat{y} - \bar{\hat{y}})\right]^2}{\sum (y - \bar{y})^2 \times \sum (\hat{y} - \bar{\hat{y}})^2} \quad (3)$$

$$r_0^2 = 1 - \frac{\sum (y - k \times \hat{y})^2}{\sum (y - \bar{y})^2} \quad (4)$$

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (5)$$

where, y and $\hat{y}$ denote the actual and predicted value of pEC50, respectively, while $\bar{y}$ and $\bar{\hat{y}}$ denote the mean of the actual and predicted value, respectively. Then, we evaluate the acceptability of the model by considering the following criteria [33–35]:

$$R^2 > 0.60$$

$$0.85 \leq k \leq 1.15$$

$$\frac{(r^2 - r_0^2)}{r^2} < 0.10$$

$$r_m^2 > 0.50$$

We also evaluate the applicability domain (AD) of the model

to investigate the limitation of the model against the data set [36,37]. The AD is calculated by using leverage approach, in which the leverage values of each compound is formulated as:

$$H = X(X^T X)^{-1} X^T \qquad (6)$$

where H is an $n \times n$ matrix that orthogonally projects vectors into space spanned by the columns of X. Then, we construct the AD region by plotting the leverage value against the standard residual value referred to as the Williams plot. The AD region lies in a square area within ± 3 for standard residual values at the y-axis and the leverage threshold (h^*) at x-axis. The leverage threshold is formulated as $h^* = 3(p+1)/n$, where p is the number of descriptor and n is the number of compounds [33,38].

2.2. Molecular docking simulation

In order to find the binding site of selected thiazolyl-pyrazoline derivatives (ligand molecules) obtained from the QSAR model into the N-terminal of protein kinase G of *Plasmodium falciparum* (receptor), molecular docking simulation is performed by using AutoDock v 4.2 packages developed by Morris and co-workers [39]. The tertiary structure of the receptor is obtained from protein data bank (PDB ID: 5E16) with resolution 1.65 Å as shown in Figure 1 [2]. The receptor is prepared by adding hydrogen atoms, and Kollman's united atom charges are given to generate the PDBQT format. Meanwhile, the selected ligands are prepared by adding gasteiger charges and polar hydrogens. The torsion atoms of ligands are identified, then PDBQT formated are created. Autogrid option is decided around the binding site of protein kinase G with enough space for the ligand rotation and translation. The grid box size is set on $52 \times 46 \times 52$ points with a grid spacing of 0.375 Å. The distance-dependent function of dielectric constant at -0.1465 is applied for the estimation of the energetic map. All of those docking parameters are prepared by using AutoDock Tools 1.5.6 [39].

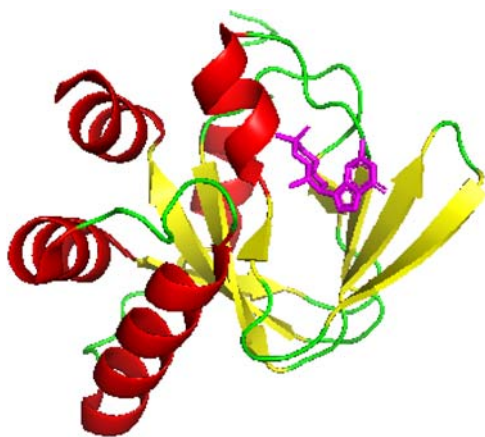The docking simulation is performed by employing the Lamarckian genetic algorithm (LGA) method to search ligand

conformation poses and orientations inside the receptor's binding site. The searching parameters are decided as follows: the maximum energy evaluation number is set to 25,000,000 per run. The number of individuals in the population and the maximum number of generations is adjusted to 150 and 2700. The rate of gene mutation and the rate of the crossover are set at 0.02 and 0.8, respectively. Other parameters are followed as the default of AutoDock v 4.2. Each optimisation involves many evaluations of the scoring function in the position, orientation, and torsion coordinates of the ligand to the receptor. The values of RMSD by <2 Å are clustered, and the binding energies are ranked based on the energy score to show the representative of ligand pocket.

2.3. Molecular dynamics simulation

To validate the stability of the complex selected from molecular docking, we perform all-atom molecular dynamics simulation on protein kinase G and selected ligands. The system is inserted with TIP3P water [40] and Na^+ ions are also added to neutralise the system. To determine the force field parameters of ligand molecules, general AMBER force field (GAFF) [41] is applied to ligands. Meanwhile, the AMBER14 force field is used for the receptor [42]. The electrostatic interactions and the constrained distance of the hydrogen atom are computed by using the Particle Mesh Ewald (PME) [43] and SHAKE [44] algorithms. The switching cutoff distance is set 10 Å, and all simulations are simulated with the time step of 2 fs.

MD simulation is started by carrying out the energy minimisation on the system. Then, the temperature is gradually increased from 0 to 300 K by performing NVT-constant for 500 ps. The temperature and the pressure of the system are kept at 300 K and 1 atm using the Langevin thermostat [45] and isotropic position scaling algorithm, respectively. The system is equilibrated with the NPT ensemble for 50 ns, and the trajectory is saved each 5000 steps (10 ps). All MD simulations are performed by using Amber16 packages [46]. CPPTRAJ tool is used to analyse the trajectories of MD simulation [47].

To validate the stability of ligand in binding with the receptor, we calculate the root-mean-square deviation (RMSD) of the complex as follows:

$$RMSD(t1) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i \parallel r_i(t_1) - r_{\text{ref},i} \parallel^2 \right]^{\frac{1}{2}}, \qquad (7)$$

where m_i is the mass of atom i, N is the total number of atoms in the model complex, M is the total mass of all atoms, and r_i is the position of atom i at the time t, and $r_{\text{ref},i}$ is the positions of ith atom in the X-ray structure, respectively.

Figure 1. (Colour online) The tertiary structure of N-terminal of protein kinase G. The structures of α-helix and β-sheet are presented by red and green colours in cartoon models, respectively. Meanwhile, the structure of cyclic guanosine-3,5-monophosphate (cGMP) refers to magenta colour by a stick model.

2.4. MM-GBSA binding energy

To estimate the binding energy of receptor–ligand complex, we employ molecular mechanics-generalised Born surface area (MM-GBSA) method developed by Miller and co-workers [48]. The binding energy is calculated from the trajectories of

MD simulations according to the following equation:

$$\Delta G_{\text{binding}} = -RT \ln Ki = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \quad (8)$$

the energy term is calculated as follows:

$$G = E_{\text{vdw}} + E_{\text{ele}} + E_{\text{GB}} + E_{\text{SA}} \quad (9)$$

Where E_{vdw}, E_{ele}, E_{GB}, and E_{SA} are the van der Waals, electrostatic, general Born solvation and surface area energies, respectively.

3. Results and discussions

3.1. QSAR analysis

To characterise the descriptors of 36 compounds of thiazolyl-pyrazoline derivatives, several QSAR approaches have been applied to find the correlation between descriptors and their chemical properties. In the present letter, anti-malaria drug candidates of thiazolyl-pyrazoline derivatives using MLR, RF, and SVM approaches are reported. Three descriptors, i.e. AATS7e, ALogP, ATSC4s, provided in Table 2 are obtained from selection protocols described in subsection 2.1.1. In Figure 2(a), we show the scatter plot of experimental values versus predicted pEC50 meanwhile, the residuals of experimental pEC50 against predicted values using MLR are shown in Figure 2(b). Based on these two figures, the pEC50 values predicted by MLR are a good agreement with experimental evidence. Also, as shown in Figure 2(b), the residual values are scattered on both areas of the zero axes. It implies that the model is free of regular error. Figure 3(a,b) shows the graph of experimental pEC50 versus calculated pEC50 values and the graph of residuals versus experimental pEC50 values calculated by using RF method. The training and test set of predicted values are noticed consistently with the experimental data as shown in 3(a). Also, from Figure 3(b), most of the residual values are distributed to the zero axes. Thus, the predicted value by RF is a good correspondence with the experimental data. For the SVM model, the scatter plot of experimental values against predicted pEC50 is provided in Figure 4(a). The predicted values are reliable with the corresponding experimental evidence. Meanwhile, in Figure 4(b), the residuals for the predicted pEC50 values are plotted against the experimental pEC50. The majority of the residuals are nearly equivalent to zero. This finding suggests that the predicted values by SVM are comparable with experimental data.

The validation parameter that indicates the acceptability and validity of the models are listed in Table 3. From the table, we find that the validation parameters of all models satisfy the threshold value. This indicates that the QSAR model developed by MLR, RF and SVM is valid and acceptable. To determine the best model, we compare the value of R^2 and r_m^2 values for both train and test sets. As for the R^2 parameter, we find that the greatest R^2 value of train and test set are obtained from RF and SVM models, respectively. The R^2 value of train and test calculated using the RF model are 0.87 and 0.75, respectively, while the R^2 value of train and test set calculated using the SVM model are 0.86 and 0.87, respectively. Meanwhile, MLR method gives the worst R^2 values for both train and test sets. The R^2 value of train and test calculated by using the MLR model are 0.67 and 0.84, respectively. As for r_m^2, we find that the greatest r_m^2 value of train and test set are acquired from RF and MLR models, respectively. The r_m^2 value of train and test calculated by using RF model are 0.82 and 0.70, respectively, while R^2 value of train and test set calculated by using MLR model are 0.66 and 0.79, respectively. To determine the best model, we consider the external validation of R^2 as more priority parameters. Hence, we find that the SVM model has the best ability in predicting the value of pEC50.

The applicability domain (AD) of the models is determined by using the leverage approach. The Williams plot of leverage values against standard residual values is presented in Figure 5. We implement the standard residue and leverage threshold to construct the AD region. According to the figure, we find that all samples lies inside the AD region for all QSAR models. This confirms the applicability of those models against both train
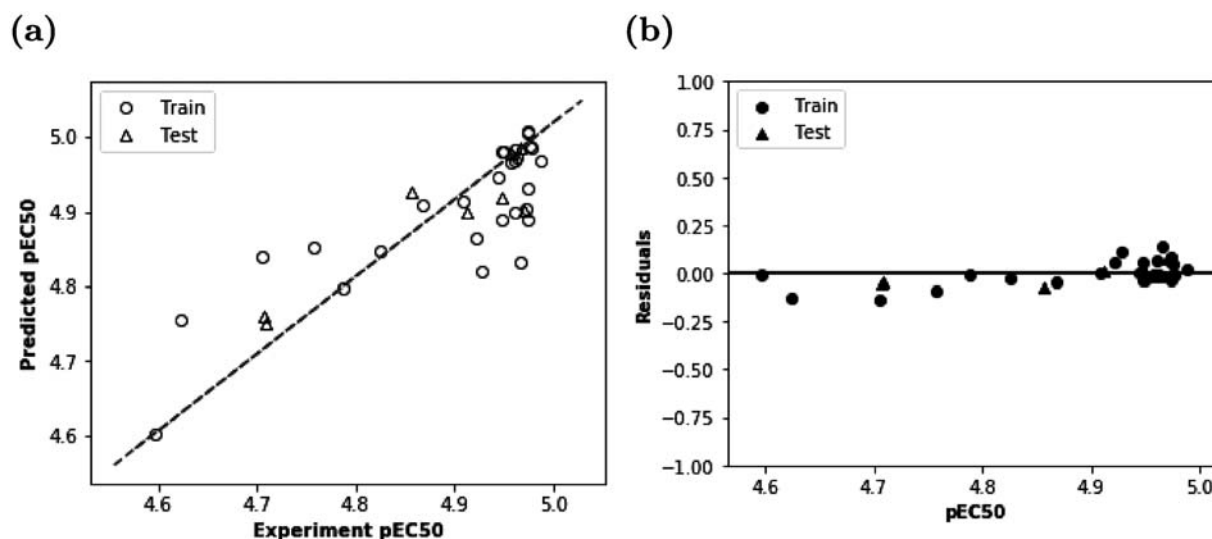

Figure 2. (a) Values of experimental pEC50 of thiazolyl-pyrazoline derivatives against the calculated values of pEC50 using MLR model. (b) Residuals of thiazolyl-pyrazoline derivatives against the experimental values of pEC50 using MLR model.
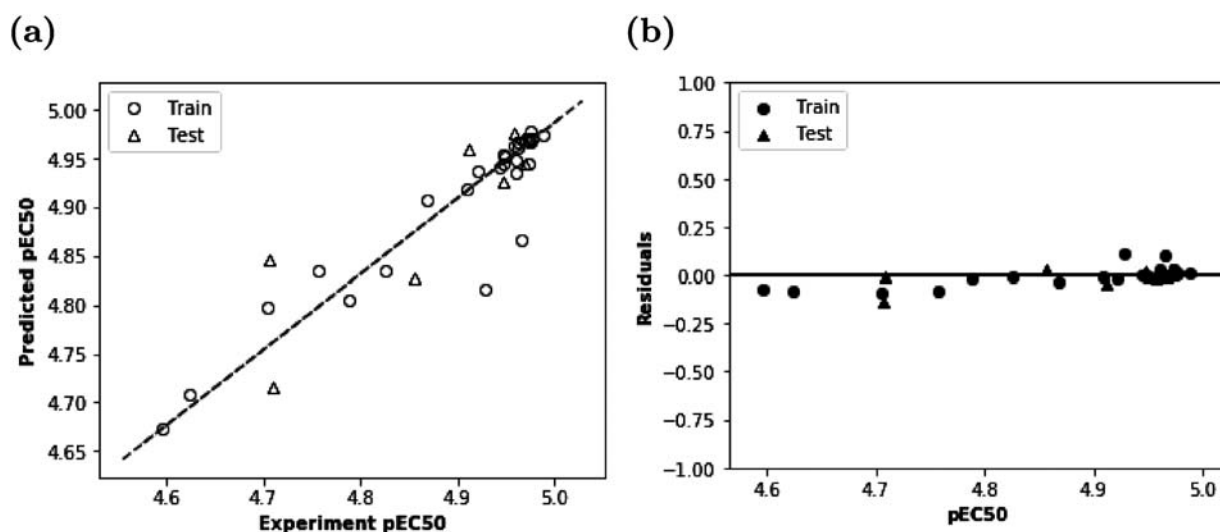
(a)

(b)

Figure 3. (a) Values of experimental pEC50 of thiazolyl-pyrazoline derivatives against the calculated values of pEC50 using RF model. (b) Residuals of thiazolyl-pyrazoline derivatives against the experimental values of pEC50 using RF model.
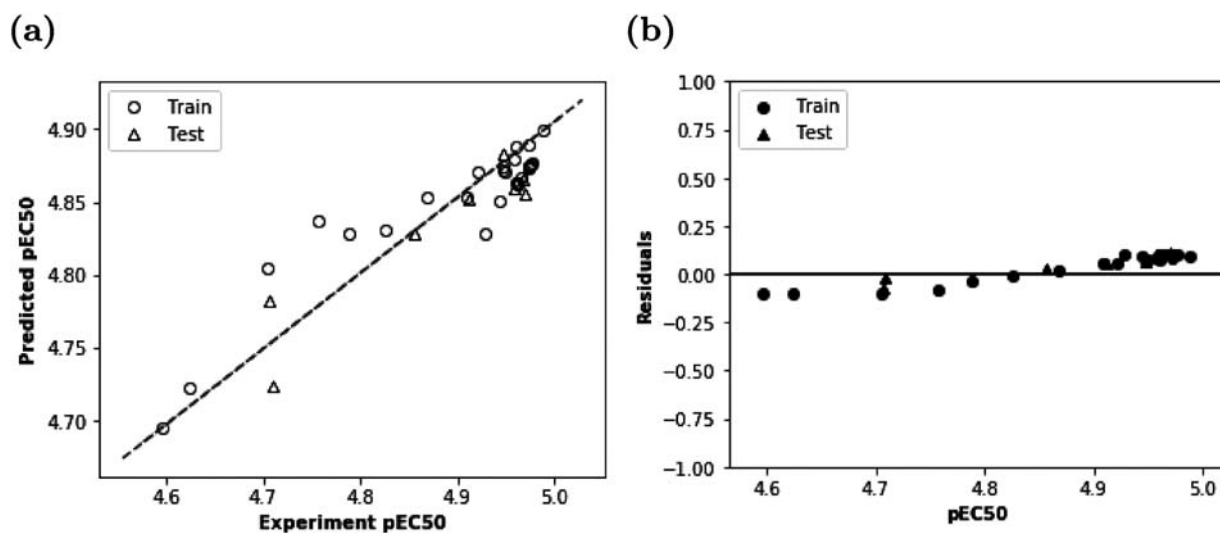
(a)

(b)

Figure 4. (a) Values of experimental pEC50 of thiazolyl-pyrazoline derivatives against the calculated values of pEC50 using SVM model. (b) Residuals of thiazolyl-pyrazoline derivatives against the experimental values of pEC50 using SVM model.

Table 3. Statistical parameters by MLR, RF and SVM models for thiazolyl-pyrazoline derivatives.

Parameter	MLR		RF		SVM		Threshold
	Train	Test	Train	Test	Train	Test	
R^2	0.67	0.84	0.87	0.75	0.86	0.87	>0.60
k	1.05	1.08	1.01	0.96	1.03	1.00	$0.85 \leq k \leq 1.15$
$\frac{(r^2 - r_0^2)}{r^2}$	0.00	0.00	0.00	0.01	0.01	0.03	< 0.10
r_m^2	0.66	0.79	0.82	0.70	0.77	0.73	>0.50

and test sets. Hence, the QSAR models developed by MLR, RF and SVM are applicable to predict pEC50 of Thiazolyl-Pyrazoline compounds.

3.2. Contribution of selected descriptors

To examine the relative importance and contribution of the descriptor with pEC50 values. We calculate the mean effect

(MF) according to the following equation:

$$MF_j = \frac{\beta_j \sum_{i}^{n} d_{ij}}{\sum_{j}^{m} \left(\beta_j \sum_{j}^{n} d_{ij} \right)} \qquad (10)$$

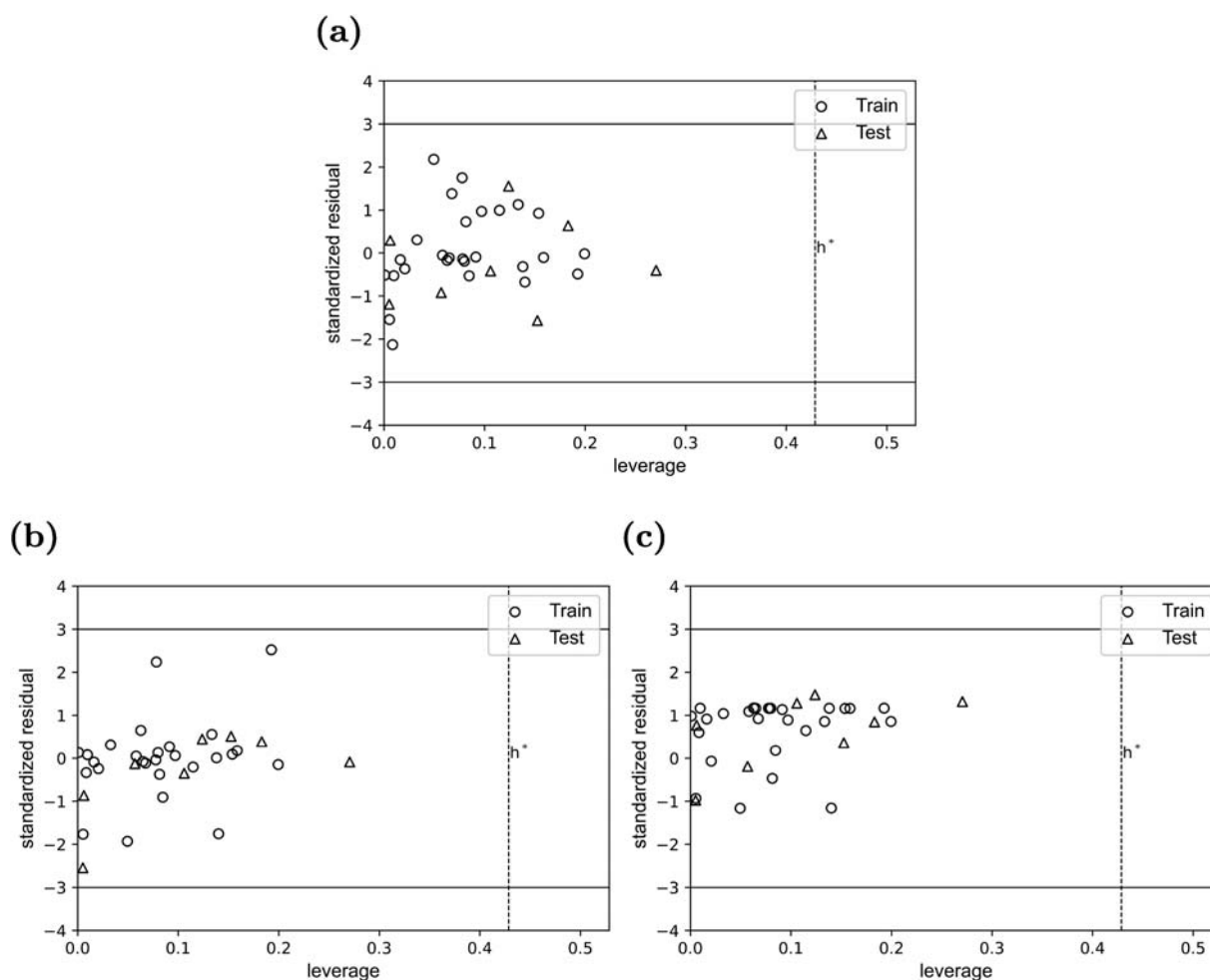MF_j and β_j represent the mean effect of each descriptor j and

Figure 5. Williams plot of applicability domain calculated by using (a) MLR, (b) RF and (c) SVM models.

the coefficient of the descriptor j, respectively. d_{ij} and m refer to the value of proposed descriptors for each compound and the number of descriptors.

Figure 6 shows MF graph of molecular descriptor for thiazolyl-pyrazoline derivatives. Three descriptors, i.e. AATS7e, ALogP, and ATSC4s, are presented. AATS7e is a molecular descriptor based on average Moreau-Broto autocorrelation of Sanderson electronegativities calculated by using physico-chemical properties such as atomic masses and covalent radius at the atomic level [49]. In MF bar, the descriptor has a positive correlation with pEC50. It assumes that the large value of AATS7e affects a high degree of anti-malaria. In Table 2, the compounds 1 (8.616) and 5 (8.20) are higher than the others. Therefore, those ligands may have better inhibitory activity than the other compounds. Another descriptor, AlogP related to the hydrophobicity of ligand, is selected in QSAR model. This descriptor describes the capability of ligand to pass the cell membrane quickly [50]. In MF estimation, the value of AlogP is positive. As a consequence, the high value of this descriptor increases the inhibitory of a compound. From our results, the compound 31 (4.941) with the highest AlogP may become more inhibitory activity than the other compounds. The last descriptor, i.e. ATSC4s, is an almost similar parameter of AATS7e based on the Moreau-Broto autocorrelation function. From MF
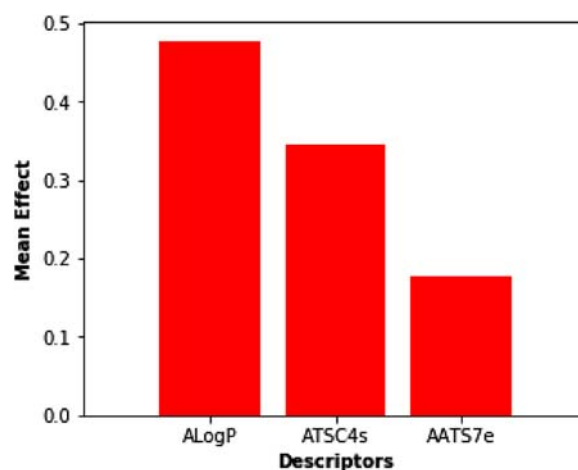

Figure 6. (Colour online) Mean effects of molecular descriptors for thiazolyl-pyrazoline derivatives.

analysis, compounds 1 (6.463) have a positive value and may have a high inhibitory activity compared to other compounds. Assumed from the MF analysis in respect to the inhibitory activity pEC50 of ligands, three compounds 1, 5, and 31 are selected for evaluating their possibilities in binding with protein kinase G by molecular docking.

3.3. Molecular docking

In molecular docking simulation, the selected configurations from the docking result are required to determine the theoretical accuracy of the obtained complex structure between ligand and receptor. To select the conformations from simulation, two strategies are considered to find the promising structure of the complex. These strategies are included selecting the lowest binding energy and concerning the binding site of the ligand into the moiety of the receptor (binding pose). The best-docked structure is found from the lowest binding energy, which presents the stable state of the ligand in binding to the site of the receptor. On the other hand, the binding pose also becomes a crucial point to screen the conformations obtained by molecular docking. The binding pose of complex provides the information of ligand position in attaching to the amino acid residues of the receptor. The best pose is selected when the ligand participates in the catalytic site (active site) of the receptor. Therefore, both the binding energy and binding pose of the ligand/receptor complex are noticed for selecting the promising conformation from the docking simulation [39,51,52].

In the previous subsection, we have performed QSAR modelling to screen 36 structures of thiazolyl-pyrazoline derivatives. From our evaluation, we obtain three compounds, including compounds 1, 5, and 31, as suggesting potent drugs against malaria. The structural formula of these compounds is shown in Figure 7. In order to expand a deeper understanding of the QSAR results. Molecular docking is performed for those compounds to find the stability complex in binding with the receptor of protein kinase G. The stable structure is achieved when the complex has negative binding energies and low inhibition constants. From our simulations, three complexes between ligand and receptor are obtained. These complexes containing compounds 1, 5, and 31 are written to become models 1, 2, and 3, respectively. In Table 4, we

summarise the thermodynamics quantities of docking results such as binding energy, inhibition constant, intermolecular energy, Van der Waals, hydrogen bond, and dehydration energies, electrostatic energy, total internal energy, torsional free energy, and unbound energy. All models have negatives binding energy. It indicates that the ligand can make a binding to the site of the receptor. Although all models have slightly different binding energies, the inhibition constant of model 2 is lower than other models. We can assume that model 2 may easily bind to the receptor for making a stable complex.

Let us discuss the interaction of ligand into the moiety of the receptor. Figure 8 shows the binding site and orientation pose of all complexes. We observe all ligands participate in hydrogen bond and hydrophobic interactions with the receptor. The details of hydrogen bonds for those complexes are listed in Table 5. In model 1, the ligand makes hydrogen bonds with residues GLY122, ALA124, ALA125, ARG132 of the receptor. In model 2, the hydrogen bonds are formed with residues of AlA124, ALA125, ARG132, SER133, and ALA134. Meanwhile, in model 3, the ligand participates hydrogen bond with residue SER133. All complexes may become stable structures since the hydrogen bonds are formed between ligand and receptor. Moreover, as mentioned in the introduction section, to deactivate the receptor of protein kinase G, the candidate drugs must prevent cGMP binding with at least one residue of the receptor such as VAL105, LYS113, MET115, PHE121, GLY122, GLU123, ALA124, ARG132, SER133, and ILE136. From our model, we see all ligands bind with at least one crucial residue for activation of protein kinase G. This indicates that all ligands can inhibit the receptor, suggesting they may have inhibitory activities as anti-malaria.

The hydrophobic interactions between ligand and receptor are also crucial interactions between ligand and receptor. Thus, we analyse this interaction for all models by LigPlot
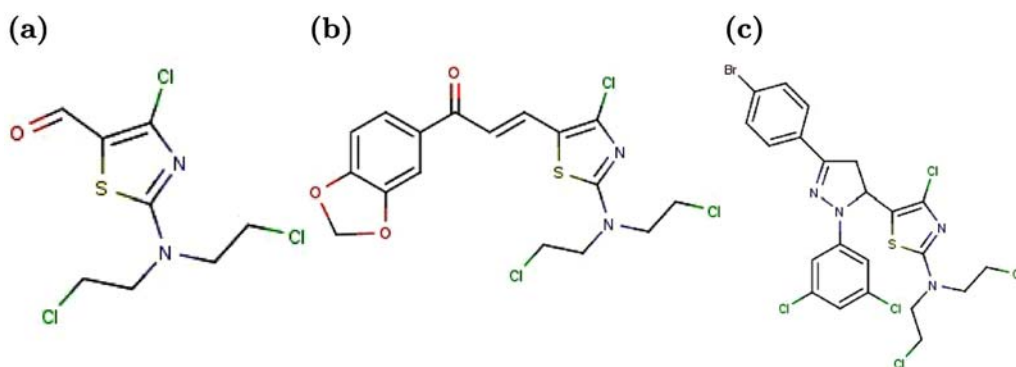

Figure 7. (Colour online) Structural formula of selected compounds from QSAR modelling (a) compound 1, (b) compound 5, and (c) compound 31.

Table 4. Thermodynamic quantities of ligand in complex with receptor obtained by molecular docking.

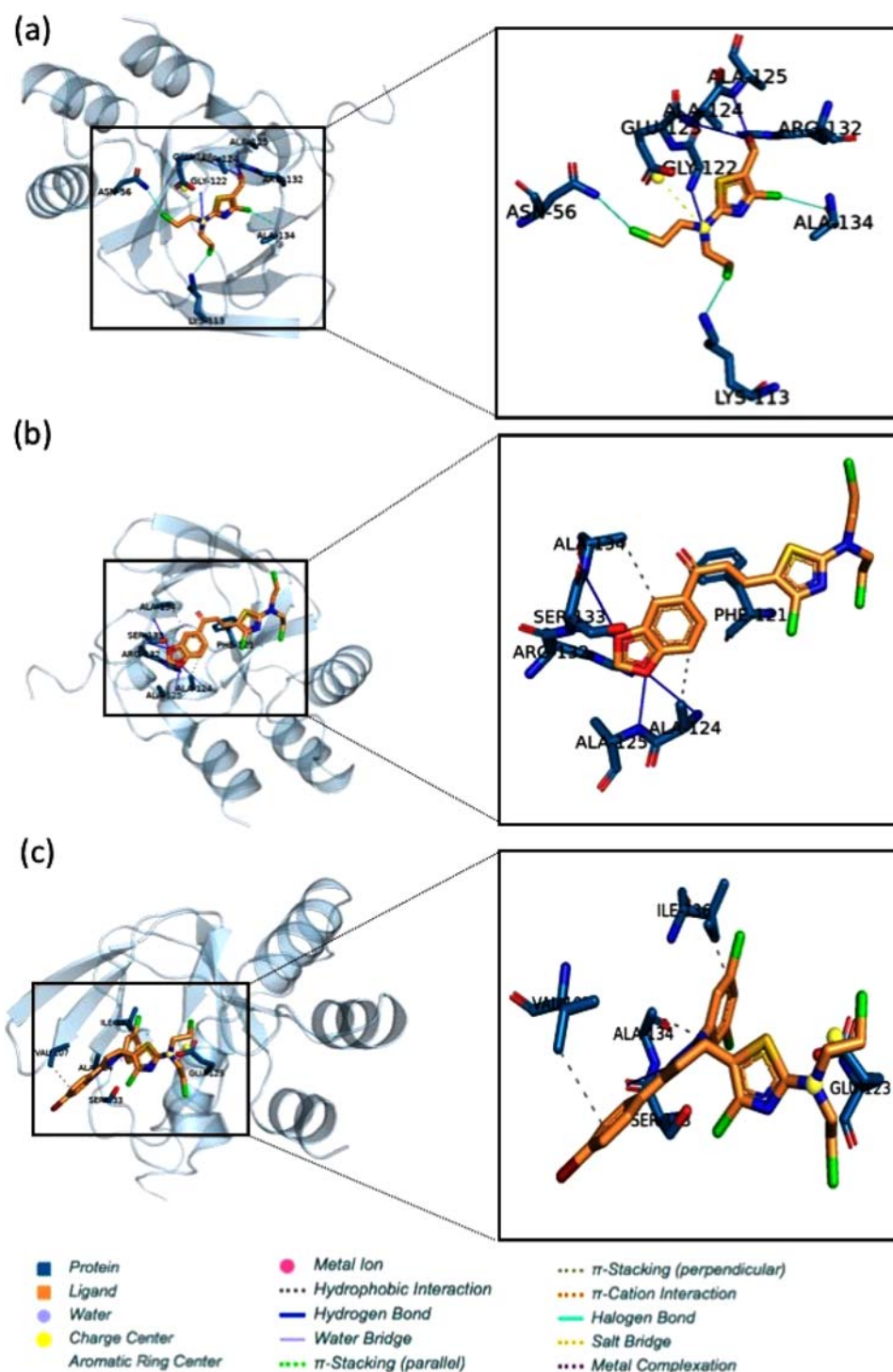| Model	Binding energy kcal/mol	Inhibition constant (Ki)/μM	Intermolecular energy (kcal/mol)	vdW + Hbond + Dsolv energy (kcal/mol)	Electrostatic energy (kcal/mol)	Internal energy (kcal/mol)	Torsional free energy (kcal/mol)	Unbound energy (kcal/mol)
1	−4.04	1110	−5.83	−5.54	−0.29	−0.42	1.79	0.42
2	−4.46	537.17	−6.85	−6.70	−0.15	−0.98	2.39	0.98
3	−4.21	825.91	−6.59	−6.51	−0.08	−1.83	2.39	−1.83

Figure 8. (Colour online) Orientation pose of ligand to the receptor. (a) Model 1, (b) Model 2, and (c) Model 3. The three dimension (3D) of complex pose is visualised by PLIP program [54] combined with Pymol v 2.3 program packages [55].

Table 5. Hydrogen bonds of thiazolyl-pyrazoline derivatives in complex with receptor.

Model	Residue	AA	Distance H-A (Å)	Distance D-A (Å)	Donor Angle	Donor Atom	Acceptor Atom
Model 1	122	GLY	2.94	3.89	153.63	1652 (N)	2244 (N3)
	124	ALA	3.57	3.91	102.05	1674 (N)	2252 (O2)
	125	ALA	2.15	3.09	152.27	1684 (N)	2252 (O2)
	132	ARG	1.91	2.86	153.80	1826 (N)	2252 (O2)
Model 2	124	ALA	2.93	3.30	102.35	1674 (N)	2243 (O3)
	125	ALA	1.88	2.80	147.86	1684 (N)	2243 (O3)
	132	ARG	2.77	3.68	148.05	1826 (N)	2243 (O3)
	133	SER	1.81	2.79	159.51	1832 (N)	2244 (O3)
	134	ALA	3.13	3.90	133.03	1843 (N)	2244 (O3)
Model 3	133	SER	2.95	3.89	154.12	2245 (N)	1837 (O2)

v.4.5.3 packages developed by Wallace and co-workers [53]. Two dimensions (2D) of the hydrophobic interactions for all complexes are shown in Figure 9. In model 1, the ligand participates in hydrophobic interaction with the residues LYS113, VAL105, ILE136, PHe121, SER120, ASN56, GLU123, ALA134, GLY122, and ILE86, ALA124, ALA125,

Figure 9. (Colour online) Hydrophobic interactions of the ligand in complex with receptor are presented by the dashed red line. (a) Model 1, (b) Model 2, and (c) Model 3.

and ARG132. In model 2, hydrophobic interactions are formed between ligand and residues SER119, THR114, MET115, SER120, LYS113, GLU123, PHE121, VAL105, ALE136, GLY122, ARG132, ALA124, SER133, ALA125. In addition, model 3 also shows hydrophobic interaction by residues of LEU57, VAL158, ASN56, GLU123, ILE127, MET115, ILE136, ALA124, VAL105, PHE121, HIS128, ALA134, SER133, and VaL107 with ligand molecules. We suppose hydrophobic interaction is also contributed to the stability of ligand in the site of the receptor.

3.4. MD analysis

From the results of molecular docking simulation provided in Figure 8, we have predicted that the promising drugs in three models may become potential inhibitors for the treatment of malaria due to their binding site to the catalytic site of the receptor. To validate the structural stability of those models in water solvent, all-atom MD simulation is performed [41,46,56]. Because of the different ligands' structures, we determine the box sizes, the number of counter ion and water molecules for each model as listed in Table 6. For the MD procedure of all models, we have presented in subsection 2.1.

In order to validate the structural stability of the complex, we analyse several parameters, i.e. root-mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration (Rg) for all models. In Figure 10, we show the RMSD value as a function of MD time for each model. The RMSD is calculated from the trajectories of MD simulations according to the equation (6). As shown in Figure 10, models 1, 2, and 3 reach the equilibrium state after 10 ns. This finding implies that our complexes are stable during the simulation. In model 1, little fluctuations are observed around 30–34 ns, but after this time, the RMSD becomes stable until the end of the simulation. This fluctuation can occur because the interactions of each residue, which includes electrostatic interaction, hydrogen bond, hydrophobic interactions or even water molecules, participate in structural rearrangement at the interface of protein. The flexibility of each model determined by estimating the RMSF factor is shown in Figure 11. As shown in the figure, all models do not significantly change the structure during the simulation. Although complex 2 looks at a higher fluctuation than models 1 and 2, all models' graphs tend to resemble similar shapes. This indicates that all complexes keep a stable along simulation. Furthermore, another parameter in relation to the protein stability is the change of protein size corresponding to the radius of gyration, as shown in Figure 12. In model 3, we find slight fluctuations around 37–42 ns. After that evolution time, the graph looks

stable again. Also, from Figure 12, we do not see any significant change of radius of gyration for each complex, indicating all complexes are stable during the simulation. Thus, assumed from all results of those stability parameters, we suggest that the complex corresponding to models 1, 2, and 3, consisting of ligand and receptor are relatively stable along the simulation.
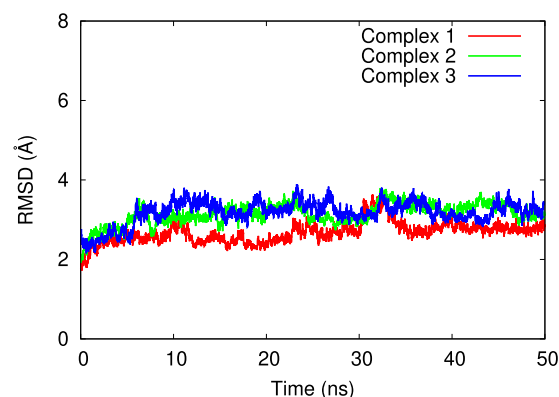

Figure 10. (Colour online) The RMSD value of complex of model 1, model 2, and model 3 is presented by red, green, and blue line colours.
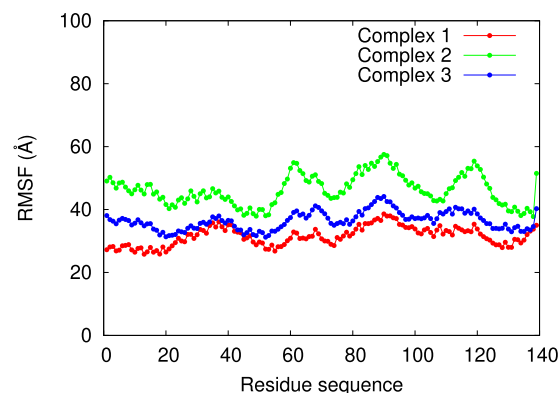

Figure 11. (Colour online) The RMSF value of complex of model 1, model 2, and model 3 is presented by red, green, and blue lines points colours.
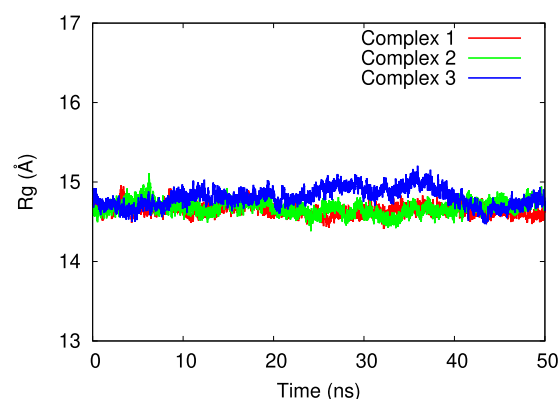

Figure 12. (Colour online) The Rg value of complex of model 1, model 2, and model 3 is presented by red, green, and blue line colours.

Table 6. Properties of simulation for all models.

Models	Identities			
	Box size (Å)	No. of ions (Na$^+$)	No. of water	Total atom
Model 1	58.924 × 65.274× 56.877	4	5744	19433
Model 2	60.156 × 65.274× 56.877	4	5935	20021
Model 3	58.924 × 65.274 × 56.877	5	5727	19409

Table 7. The binding energy and the contribution of each energy term. The units for all contributions are given by kcal/mol.

Model	E_{vdw}	E_{ele}	E_{GB}	E_{SA}	ΔG_{MM}	ΔG_{Solv}^{GBSA}	ΔG_{bind}
Model 1	−23.44	−12.56	27.91	−3.19	−36.01	24.72	−11.28
Model 2	−25.26	−27.63	38.24	−3.49	−52.89	34.74	−18.14
Model 3	−26.02	−75.26	88.17	−3.29	−101.29	84.88	−16.41

To decide which model are the most stable structure in the water solvent, the binding energy using MM-GBSA method is calculated according to the equation (9). The MD trajectories are selected from the equilibrium state from 20–50 ns. Thus, 3000 frames are used to calculate the binding energy. From our estimation, the binding free energies, including the contribution of energy for each model, are summarised in Table 7. We find that model 2 has the lowest binding energy at −18.14 kcal/mol. For models 1 and 2, the binding energy is obtained at −16.41 kcal/mol and −11.28 kcal/mol, respectively. This finding implies that the most stable complex may be found in model 2. Also, the binding energy from MM-GBSA method is a good correspondence with our result obtained by molecular docking where the lowest binding energy by these two methods is found in model 2, then followed by models 3 and 1, respectively. Further, as shown in Table 7, we see the E_{ele} and E_{Gb} values for all models have significantly different values. This indicates those energy terms influence the difference of binding energy for each model.

Since model 2 becomes the most stable structure obtained by the estimation of the binding energy using MM-GBSA method. We extract two snapshot conformations of the complex at 30 ns (blue) and 50 ns (magenta) of MD simulations, then aligned to the structure obtained from molecular docking (yellow). The structural superimposes for those molecular docking and MD snapshots are shown in Figure 13. From this figure, we can see the complex fluctuation in respect to the conformational changes of receptor and ligand molecule. For the receptor molecule, no significant fluctuations are observed for both docking and MD structures. This finding reveals that the protein keeps stable in the system during the simulations. Meanwhile, in ligand molecule, the structure involves the large fluctuations shown in Figure 13. These fluctuations may increase the probability of ligand in binding with the receptor to form a complex structure. Hence, the ligand of model 2 may become the most promising drug against malaria not only have the lowest binding energy but also participate in interactions easily to the catalytic site of protein kinase G.
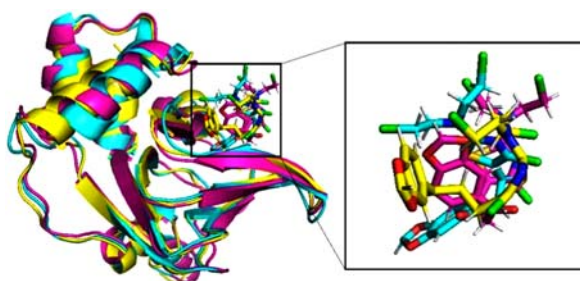

Figure 13. (Colour online) Superimposes of the complex of model 2 obtained from docking (yellow) and MD simulations at 30 ns (blue) and 50 ns (magenta) are presented by cartoon model.

4. Summary

Inhibition of protein kinase G becomes a crucial point for the treatment of malaria due to its essential role in *Plasmodium* parasite life. In this present study, the purpose of searching novel, potent and selective protein kinase G inhibitor as anti-malaria is performed by combining several computational approaches such as QSAR modelling, molecular docking, and all-atom MD simulation. From QSAR modelling, we employ MLR, RF, and SVM approaches to find the relation between the inhibitory action, i.e. pEC50, and the variable of various descriptors. The several validation parameters of each model are estimated to validate the strong/weak correlation of predicted and experiment value of pEC50. Also, we estimate the mean effect to examine the relative importance and contribution of the descriptors with pEC50 and select three potent drugs for the anti-malaria agent. Further, those selected drugs are used as ligand molecules for molecular docking. From our results, we predict there complex of models 1, 2, and 3 bind to the catalytic site of protein kinase G. This finding implies those ligands may become inhibitors for *Plasmodium falciparum*. To validate the stability of those complexes, all-atom MD simulation is performed on all models. Based on the parameters of RMSD, RMSF, and Rg, all complexes are stable along simulation since no significant fluctuations are observed in those parameters. Besides, the binding energy for complexes is estimated by MM-GBSA to decide the most stable structure in the water solvent. From our calculation, model 2 may become the most stable structure based on the binding energy score. Therefore, the snapshot structures of model 2 from molecular docking and MD simulations are extracted to see the conformational changes of the complex. The side chain of the ligand becomes large fluctuations during the simulation. We imply that the fluctuation of the ligand may increase the probability of binding with the receptor of protein kinase G. Finally, assumed from our simulations, three ligands may have potential as inhibitors of protein kinase G and the ligand of model 2 may become the promising drug against malaria. Therefore, for further safety and health interest, the suggested drugs of our computational assessment need further in vitro and in vivo analysis or event preclinical trials.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Arwansyah Arwansyah http://orcid.org/0000-0003-1784-5942
Abdur Rahman Arif http://orcid.org/0000-0001-8722-2949
Gita Syahputra http://orcid.org/0000-0001-6164-9397
Isman Kurniawan http://orcid.org/0000-0003-3485-3063

References

[1] Who. World malaria report. World Malaria Report. 2020.

[2] El Bakkouri M, Kouidmi I, Wernimont AK, et al. Structures of the cgmp-dependent protein kinase in malaria parasites reveal a unique structural relay mechanism for activation. Proc Natl Acad Sci. 2019;116(28):14164–14173. Available from: https://www.pnas.org/content/116/28/14164.

[3] Vidadala RSR, Ojo KK, Johnson SM, et al. Development of potent and selective *Plasmodium falciparum* calcium-dependent protein kinase 4 (pfcdpk4) inhibitors that block the transmission of malaria to mosquitoes. Eur J Med Chem. 2014;74:562–573. Available from: http://www.sciencedirect.com/science/article/pii/S0223523414000105.

[4] Moussaoui D, Robblee JP, Auguin D, et al. Full-length *Plasmodium falciparum* myosin a and essential light chain pfelc structures provide new anti-malarial targets. eLife. 2020;9:1–24.

[5] Tuteja R. Helicases feasible antimalarial drug target for *Plasmodium falciparum*. FEBS J. 2007;274(18):4699–4704. DOI:10.1111/j.1742-4658.2007.06000.x.

[6] Daniyan MO, Blatch GL. Plasmodial Hsp40s: new avenues for antimalarial drug discovery. Curr Pharm Des. 2017; 23(30):4555–4570. DOI:10.2174/1381612823666170124142439.

[7] Doerig C, Billker O, Pratt D, et al. Protein kinases as targets for antimalarial intervention: kinomics, structure-based design, transmission-blockade, and targeting host cell enzymes. Biochim Biophys Acta (BBA) Proteins Proteom. 2005;1754(1):132–150. Inhibitors of Protein Kinases. 4th International Conference, Inhibitors of Protein Kinases and Associated Workshop: Modelling of Specific Molecular Recognition Processes (Warsaw, Poland, June 2005. p. 25–29); Available from: https://www.sciencedirect.com/science/article/pii/S1570963905003377.

[8] Ojha PK, Roy K. Exploring molecular docking and QSAR studies of plasmepsin-II inhibitor di-tertiary amines as potential antimalarial compounds. Mol Simul. 2011;37(9):779–803. DOI:10.1080/08927022.2010.548384.

[9] Ibrahim MA, Abdelrahman AH, Hassan AM. Identification of novel *Plasmodium falciparum* pi4kb inhibitors as potential anti-malarial drugs: homology modeling, molecular docking and molecular dynamics simulations. Comput Biol Chem. 2019;80:79–89. Available from: http://www.sciencedirect.com/science/article/pii/S1476927118302950.

[10] Lima MNN, Cassiano GC, Tomaz KCP, et al. Integrative multi-kinase approach for the identification of potent antiplasmodial hits. Front Chem. 2019;7:773. Available from: https://www.frontiersin.org/article/10.3389/fchem.2019.00773.

[11] Abdullahi M, Shallangwa GA, Uzairu A. In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of h5n1 influenza a virus subtype Beni-Suef University. J Basic Appl Sci. 2020 Jan;9(1):2. DOI:10.1186/s43088-019-0023-y.

[12] Kurniawan I, Rosalinda M, Ikhsan N. Implementation of ensemble methods on QSAR study of ns3 inhibitor activity as anti-dengue agent. SAR QSAR Environ Res. 2020;31(6):477–492. DOI:10.1080/1062936X.2020.1773534.

[13] Kurniawan I, Tarwidi D, Jondri J. QSAR modeling of PTP1b inhibitor by using genetic algorithm-neural network methods. J Phys Conf Ser. 2019 Mar;1192:012059. DOI:10.1088/1742-6596/1192/1/012059.

[14] Tadayon M, Garkani-Nejad Z. In silico study combining QSAR, docking and molecular dynamics simulation on 2,4-disubstituted pyridopyrimidine derivatives. J Recep Signal Transduct. 2019;39(2):167–174. DOI:10.1080/10799893.2019.1641821.

[15] Hdoufane I, Stoycheva J, Tadjer A. Qsar and molecular docking studies of indole-based analogs as HIV-1 attachment inhibitors. J Mol Struct. 2019;1193:429–443. Available from: http://www.sciencedirect.com/science/article/pii/S0022286019306209.

[16] Azmi HF, Lhaksmana KM, Kurniawan I. QSAR study of fusidic acid derivative as anti-malaria agents by using artificial neural network-genetic algorithm. In: 2020 8th International Conference on Information and Communication Technology (ICoICT); Jun 2020. p. 1–4.

[17] Rahman F, Lhaksmana KM, Kurniawan I. Implementation of simulated annealing-support vector machine on QSAR study of fusidic acid derivatives as anti-malarial agent. In: 2020 6th International Conference on Interactive Digital Media (ICIDM); Dec 2020. p. 1–4.

[18] Kurniawan I, Fareza MS, Iswanto P. CoMFA, molecular docking and molecular dynamics studies on cycloguanil analogues as potent antimalarial agents. Indones J Chem. 2020 Sep;21(1):66–76. Number: 1; Available from: https://jurnal.ugm.ac.id/ijc/article/view/52388.

[19] Blaney JM, Dixon JS. A good ligand is hard to find: automated docking methods. Perspect Drug Discov Des. 1993 Dec;1(2):301–319. DOI:10.1007/BF02174531.

[20] Baxter CA, Murray CW, Waszkowycz B, et al. New approach to molecular docking and its application to virtual screening of chemical databases. J Chem Inf Comput Sci. 2000;40(2):254–262. DOI:10.1021/ci990440d.

[21] Kitchen DB, Decornez H, Furr JR, et al. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004 Nov;3(11):935–949. DOI:10.1038/nrd1549.

[22] Cuartas V, Robledo SM, Vlez ID, et al. New thiazolyl-pyrazoline derivatives bearing nitrogen mustard as potential antimicrobial and antiprotozoal agents. Arch Pharm (Weinheim). 2020;353(5): e1900351. DOI:10.1002/ardp.201900351.

[23] Yap CW. Padel-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011;32 (7):1466–1474. DOI:10.1002/jcc.21707.

[24] Kim S, Cho KH. Pyqsar: a fast QSAR modeling platform using machine learning and jupyter notebook. Bull Korean Chem Soc. 2019;40(1):39–44. DOI:10.1002/bkcs.11638.

[25] Breiman L, Friedman JH, Olshen RA, et al. Classification and regression trees. New York (NY): Chapman and Hall/CRC; 1984.

[26] Breiman L. Random forests. Mach Learn. 2004 Aug;45(3):5–32.

[27] Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995 Sep;20(3):273–297. DOI:10.1007/BF00994018.

[28] Vapnik VN. Statistical learning theory. United States: Wiley-Interscience; 1998.

[29] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput. 2004 Aug;14(3):199–222. DOI:10.1023/B:STCO.0000035301.49549.88.

[30] Rossum G. Python reference manual. NLD; 1995.

[31] Roy K, Chakraborty P, Mitra I, et al. Some case studies on application of 'rm2' metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. J Comput Chem. 2013;34(12):1071–1082. DOI:10.1002/jcc.23231.

[32] Gajo GC, De Assis TM, Assis LC, et al. Quantitative structure-activity relationship studies for potential rho-associated protein kinase inhibitors. J Chem. 2016: 9198582. DOI:10.1155/2016/9198582.

[33] Roy K, Mitra I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Comb Chem High Throughput Screen. 2011 Jul;14 (6):450–474.

[34] Veerasamy R, Rajak H, Jain A, et al. Validation of QSAR models – strategies and importance. Int J Drug Design and Discov. 2011 Jul;2:511–519.

[35] Ojha PK, Mitra I, Das RN, et al. Further exploring rm2 metrics for validation of QSPR models. Chemometr Intel Laborat Syst. 2011 May;107(1):194–205. Available from: https://www.sciencedirect.com/science/article/pii/S016974391100061X.

[36] Roy K, Kar S, Ambure P. On a simple approach for determining applicability domain of QSAR models. 2015;145:22–29. DOI:10. 1016/j.chemolab.2015.04.013.

[37] De Assis TM, Gajo GC, De Assis LC, et al. QSAR models guided by molecular dynamics applied to human glucokinase activators. Chem Biol Drug Des. 2016;87:455–466. DOI:10. 1111/cbdd.12683.

[38] Sahigara F, Mansouri K, Ballabio D, et al. Comparison of different approaches to define the applicability domain of QSAR models. Molecules. 2012 Apr;17(5):4791–4810.

[39] Morris GM, Huey R, Lindstrom W, et al. Autodock4 and autodock-tools4: automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec;30(16):2785–2791. Available from: https://pubmed.ncbi.nlm.nih.gov/19399780.

[40] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–935.

[41] Wang J, Wolf RM, Caldwell JW, et al. Development and testing of a general Amber force field. J Comput Chem. 2004;25(9):1157–1174.

[42] Maier JA, Martinez C, Kasavajhala K, et al. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. J Chem Theory Comput. 2015;11(8):3696–3713. DOI:10. 1021/acs.jctc.5b00255.

[43] Essmann U, Perera L, Berkowitz ML, et al. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577–8593. DOI:10. 1063/1.470117.

[44] Ryckaert JP, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23(3):327–341. Available from: http://www.sciencedirect.com/science/article/pii/0021999177900985.

[45] Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of n-acet-ylalanyl-n-methylamide. Biopolymers. 1992;32(5):523–535. DOI:10.1002/bip.360320508.

[46] Salomon-Ferrer R, Case DA, Walker RC. An overview of the amber biomolecular simulation package. WIREs Comput Mol Sci. 2013;3 (2):198–210.

[47] Roe DR, Cheatham TE. Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9(7):3084–3095.

[48] Miller BR, Mcgee TD, Swails JM. MMPBSA. py: an efficient program for end-state free energy calculations. 2012.

[49] Todeschini R, Consonni V. Molecular descriptors for chemoinformatics; 2009.

[50] Ghose AK, Crippen GM. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships modeling dispersive and hydrophobic interactions. J Chem Inf Comput Sci. 1987;27(1):21–35. DOI:10.1021/ci00053a005.

[51] Giacoppo JO, França TC, Kuča K, et al. Molecular modeling and in vitro reactivation study between the oxime BI-6 and acetylcholinesterase inhibited by different nerve agents. J Biomol Struct Dyn. 2015;33(9):2048–2058. DOI:10.1080/07391102.2014.989408.

[52] EA de Lima W, F Pereira A, A de Castro A, et al. Lett drug des discov. flexibility in the molecular design of Acetylcholinesterase reactivators: probing representative conformations by chemometric techniques and Docking/QM calculations. 2015;13(5):360–371. DOI:10.2174/1570180812666150918191550.

[53] Wallace AC, Laskowski RA, Thornton JM. Ligplot: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng Design Select. 1995;8(2):127–34. DOI:10.1093/protein/8.2.127.

[54] Salentin S, Schreiber S, Haupt VJ, et al. PLIP: fully automated protein-ligand interaction profiler. Nucl Acids Res.2015;43 (W1):443–447. DOI:10.1093/nar/gkv315.

[55] DeLano WL. The PyMOL Molecular Graphics System, Version 2.3; 2020.

[56] Kawaguchi K, Arwansyah MS, Kataoka T, et al. Theoretical study of conformational transition of Cdk4 by association of cyclin d3. Mol Phys. 2019;117(17):2355–2361. DOI:10.1080/00268976.2018.1563725.