Indian Journal of Chemistry
Vol. 60B, May 2021, pp. 727-731

3D QSAR analysis on quinoxaline derivatives as anti-malarial using K-nearest neighbour molecular field analysis

Achal Mishra*[a], Yogesh Vaishnav[a], Arvind Kumar Jha[a] & Shekhar Verma[b]

[a] Faculty of Pharmaceutical Sciences Shri Shankaracharya Technical Campus, Bhilai 491 001, India

[b] University College of Pharmacy, Pt Deendayal Upadhyay Memorial Health Sciences and Ayush University, Raipur 493 661, India

E-mail: achal.mishra03@gmail.com

Received 20 August 2020; accepted (revised) 3 March 2021

In the present article, k nearest neighbour molecular field analysis (kNN-MFA) method was used to develop a three dimensional quantitative structure activity relationship (3D-QSAR) model. In this study 37 derivatives of quinoxaline having antimalarial activity were used. Sphere exclusion (SE) algorithm was used to create the biological activity data set in to into training and test set. For model generation kNN-MFA method has coupled with stepwise, simulated annealing and genetic algorithm this method provides various models, in which the most significant model developed by stepwise backward-forward method with predictive internal q^2=0.7589 and external predictivity (pred_r^2 = 0.4752). In the presented model electrostatic descriptors play crucial role for activity. Electrostatic descriptor (E_137) indicates regions in which electron withdrawing groups are favourable and descriptor (E_939) represents electron rich or electron donating groups are advantageous in particular region. The counter map/ plot of this model further helps to understand the relationship of structural feature of derivative of quinoxaline and its biological activity this would be applied for designing of new potent antimalarial containing quinoxaline as lead.

Keywords: 3D-QSAR, kNN-MFA, antimalarial, quinoxaline derivatives, sphere exclusion (SE) algorithm

Malaria is a life intimidating ailment originated by means of parasites and later on transmitted to individuals through infected female anopheles mosquitoes bites. This infection has globally important as in many regions of the world it is a cause of mortality and morbidity and creates social and economic loss to those regions. WHO report of 2019 clearly shows that in 2018 around 228 million causes are reported worldwide and 405000 deaths from this. Approximately 47% *P. vivax* cases are reported in India in total 53% of globally. Most of the drugs present in the market to treat this infection now are less active or not active against most of the malarial parasites because of resistance developed by parasites, so there is an urgent need to develop new molecules to restrict the development and existence of this infection[1].

Quantitative structure-activity relationships (QSAR) is a hypothetical method which can helps to correlate the biological properties or activates with descriptors that should be mathematically calculated and this relationship is given by some mathematical equations[2].

Some quinoxaline derivatives show good activity against some malarial parasites although various QSAR studies were reported and gives various models for antimalarial drug designing but nitrogen containing quinoxaline heterocyclic compounds are not properly checked for QSAR model development. In the present work a 3D-QSAR has to be performed for QSAR model development and try to stabilize a good correlation between biological activities *vs*. descriptors. This will help to established new molecules with less or no resistance[3-6] (Table I).

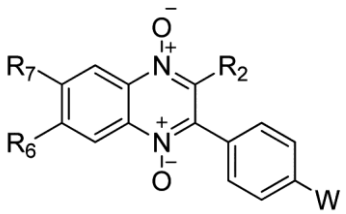
Result and Discussion

Model - 3D-RAN-70%-SWFB-kNN-trial 7

A total of ten trials were done to establish best 3D model. The above model was considered as best in view of q^2 and predicted r^2 values (Figure 1 and Figure 2). Both the values were in optimum range which made this model as best model and the results are shown in Table II.

Model X was generated by means of random selection method followed by 70% in which the whole set of molecules were separated into training and test set. In this model, 25 molecules were considered as training set and rest of the 12 molecules were considered as test set. From the generated statistical equation, it was confirmed that the model

Table I — Data set of 37 molecules of quinoxaline derivatives and their biological activities in the form of IC_{50} converted in –log IC_{50}



S No	ID	R2	W	R6	R7	IC_{50} (μM)	-log IC_{50}
1	1	CN	H	H	H	2.35± 0.04	8.6289
2	3	CN	H	H	CH3	1.55±0.76	8.8096
3	4	CN	H	H	OCH3	1.47±0.24	8.8326
4	5	CN	Cl	H	H	1.01±0.37	8.9956
5	7	CN	Cl	H	CH3	0.48±0.13	9.3187
6	8	CN	Cl	H	OCH3	0.73±0.03	9.1366
7	9	CN	CH3	H	H	3.46±1.08	8.4609
8	11	CN	CH3	H	CH3	1.30±0.55	8.886
9	12	CN	CH3	H	OCH3	0.81±0.13	9.0915
10	13	CN	COOCH3	H	H	6.54±1.03	8.1844
11	15	CN	COOCH3	H	CH3	2.15±0.80	8.6675
12	16	CN	COOCH3	H	OCH3	3.16±0.45	8.5003
13	17	CN	F	H	H	3.91±0.92	8.4078
14	18	CN	F	H	Cl	5.83±0.60	8.2343
15	19	CN	F	H	CH3	0.88±0.13	9.0555
16	20	CN	F	H	OCH3	0.87±0.19	9.0604
17	22	CN	F	CH3	CH3	1.26±0.16	8.8996
18	23	CN	OCH3	H	H	3.14±0.44	8.503
19	25	CN	OCH3	H	CH3	2.77±0.65	8.5575
20	26	CN	OCH3	H	OCH3	2.29±0.12	8.6401
21	28	CN	OCH3	CH3	CH3	3.02±0.31	8.5199
22	29	CN	OCF3	H	H	1.27±0.17	8.8961
23	30	CN	OCF3	H	Cl	0.71±0.34	9.1487
24	31	CN	OCF3	H	CH3	0.75±0.28	9.1249
25	32	CN	OCF3	H	OCH3	0.93±0.48	9.0315
26	36	CN	F	H	F	5.78±0.20	8.238
27	37	CN	Cl	H	F	4.21±0.57	8.3757
28	40	CN	OCF3	H	F	1.70±0.19	8.7695
29	42	CN	H	H	CF3	0.33±0.12	9.4814
30	43	CN	F	H	CF3	0.66±0.03	9.1804
31	44	CN	Cl	H	CF3	1.56±0.19	8.8068
32	45	CN	CH3	H	CF3	0.72±0.09	9.1426
33	46	CN	OCH3	H	CF3	2.44±0.30	8.6126
34	47	CN	OCF3	H	CF3	1.93±0.31	8.7144
35	48	CN	COOCH3	H	CF3	2.03±0.62	8.6925
36	54	COOC2H5	H	H	Cl	3.22±0.34	8.4921
37	59	COOC2H5	H	CH3	CH3	1.18±0.05	8.9281

has good projecting ability of 75%. The electrostatic descriptor plays a significant role in determining biotic activity. Two different electrostatic descriptors E_939 and E_137 with negative range of -1.0625 - 0.8001 and positive range of 0.0073, 0.1361 respectively. The contribution of respective relative electrostatic fields indicates that electrostatic fields were predominant which was shown in Figure 3. The actual and predicted activities of all the compounds in the training and test set are in well agreement with their respective biological activities and shown in Table III and Table IV. The actual and predicted biotic activity of training and test set was depicted in Figure 4. The involvement plot of electrostatic field's

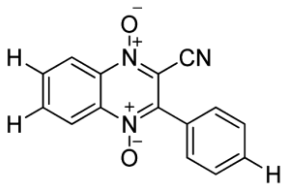


Figure 1 — Structure of template [1]

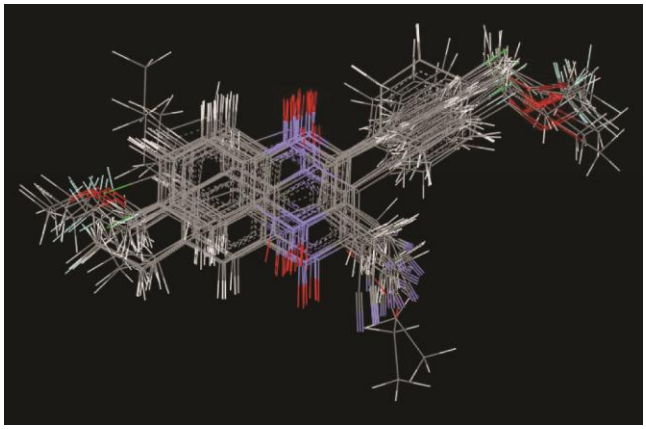


Figure 2 — Template based 3D-alignment of molecules

Table II — Results of 3D-QSAR analysis using kNN-MFA method (k nearest neighbour) by random selection 70% in connection with step wise forward backward (SWFB) as variable selection method

S No	Trial No	Test Set Molecules	RAN-SWFB-kNN
01	07	2,10,13,23,24,26,27, 29,31,35,36,37	Optimum Components = 2 n = 25 Degree of freedom = 21 q^2 = 0.7589 q^2 se = 0.1361 pred_r^2 = 0.4752 pred_r^2se = 0.2987

Optimum component = 2, n = 25, Degree of freedom = 21, q^2=0.7589, q^2_se = 0.1361, Pred_r^2 = 0.4752, pred_r^2se = 0.2987. The values of different parameters in model X was shown in Table II.

interaction point out relative regions of the local fields around the aligned molecules and foremost to activity variation in the generated model. The blue colored ball represents the electrostatic field descriptor and the positive value range (E_137 0.0073 0.1361) indicates regions in which electron withdrawing groups are favourable whereas negative value range (E_939 -1.0625 -0.8001) represents electron rich or electron donating groups are advantageous in particular region. 3D QSAR model reveals that the electrostatic descriptors with positive as well as negative coefficient values are from R_6 and R_7 position of substituted aryl ring. The fitness plot of

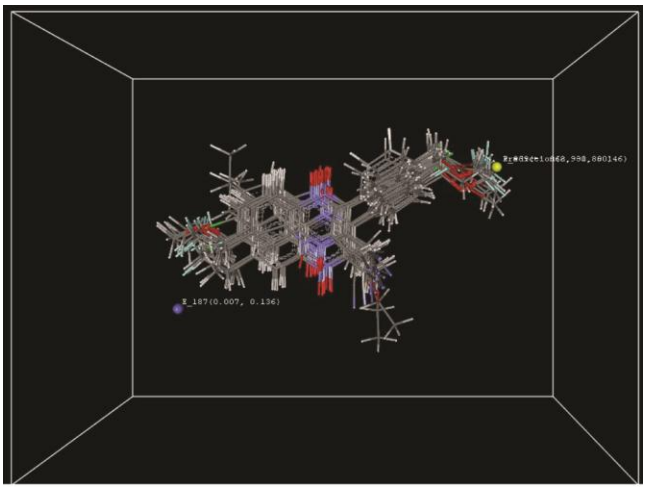


Figure 3 — Grid point of descriptors contributing in 3D QSAR model

Table III — The actual and predicted activities of the training set

Training Set	Actual Values	Predicted Values
07	8.4609	8.5932
14	8.2343	8.5449
15	9.0555	8.9476
20	8.6401	8.69505
11	8.6675	8.5449
05	9.3187	9.04
08	8.886	8.8041
25	9.0315	8.97825
33	8.6126	8.5932
22	8.8961	8.807
32	9.1426	9.088
01	8.6289	8.5642
06	9.1366	9.07595
18	8.503	8.58505
17	8.8996	9.02555
34	8.7144	8.8328
03	8.8326	8.6538
16	9.0604	9.11405
28	8.7695	8.80525
04	8.9956	8.97755
12	8.5003	8.75005
19	8.5575	8.48195
21	8.5199	8.7493
30	9.1804	9.15715
09	9.0915	9.0985

actual *versus* predicted biotic activity was depicted in Figure 5.

Materials and Methods

The present work consist of 37 derivatives of quinoxaline as this were reported to have antiplasmodial activity and the IC_{50} [inhibition constant (μM)] values were converted to the pIC_{50}

Table IV — The actual and predicted activities of the test set

Test Set	Actual Values	Predicted Values
29	9.4814	9.16144
23	9.1487	8.74537
37	8.9281	9.24915
13	8.4078	8.56629
31	8.8068	8.47796
24	9.1249	8.81068
02	8.8096	8.94093
35	8.6925	8.87152
36	8.4921	8.413
26	8.238	8.3943
27	8.3757	8.66801
10	8.1844	8.64883

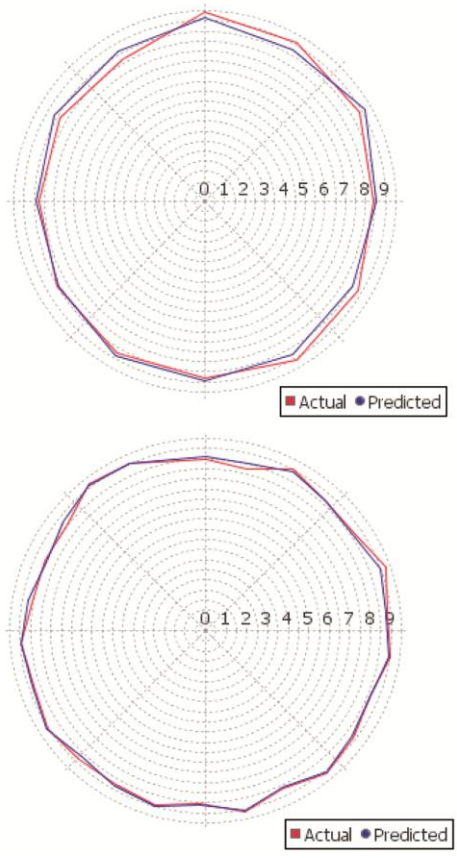


Figure 4 — Actual and Predicted biotic activity of Training and Test set molecules

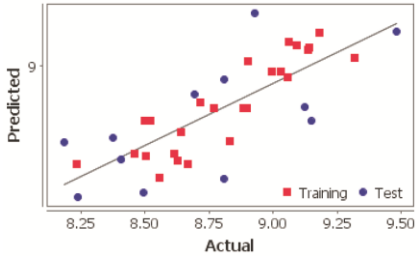


Figure 5 — Fitness plot of actual *vs* predicted biotic activity

values which is the minus logarithm of the IC_{50} [7] that has represented in Table I.

Molecular modeling study/ 3D-QSAR study

3D QSAR and kNN- MFA studies were performed in Molecular Design Suite (MDS) installed in Lenovo computer system having genuine Intel Pentium i3 processor with windows XP operating system. Structures of the compounds were drawn using the 2D draw application and convert them to 3D by convert to 3D tool. Energy minimization and geometry optimization has done by Merck Molecular Force Field (MMFF) method and maximum number of cycle has 1000, convergence criteria that is root mean square gradient selected as 0.01 and medium dielectric constant is takes 1.0 in dielectric properties for steric energy and electrostatic cutoff 30.0 and 10.0 Kcal/mol were used[7].

For alignment of dataset template based alignment method was used by choosing the most active compound as template for alignment (Table I) molecule 1 has also used as reference for alignment (Figure 1), now all the molecules were align in template and reference molecule.

The dataset were aligned by template based alignment method using most active molecule (37, Table I) as a reference molecule **1** and structure **1** as a template (Figure 1). The alignment of all the molecules on the template has shown in Figure 2. In the template based alignment method, a template structure was defined and used as a basis for alignment of a set of molecules. After alignment of all molecule a grid or lattice has set on 2 Å resolution and now various descriptors represents by electrostatic, steric and hydrophobic interplay energies were calculated by setting the value of probe of charge +1. These descriptors show how the molecule binds with active site.

The data set of all compounds was divided in to training and test set using sphere exclusion method. In this algorithm training sets were constructed by capturing whole descriptor space which is cover by representative points. Size of test and training sets were deicide by dissimilarity value represented by c, larger the value of c larger the size of test set and small the size of training set. In QSAR predication the value of dissimilarity level gives the predictive ability of the model if it is high than the predicative power may ne decreases. After the generation of training and test set k-NN method has to be applied [8, 9].

k-NN-MFA method

In k nearest neighbour molecular field investigation is based on distance learning path in which unidentified element is classified giving to the mainstream of its k-NN in the training set of molecules. A large number of models were developed any means of this k-NN methodology.

k-NN approach begins with the selection of training and test sets followed by choosing descriptors which were later created over grid. The grid frame points show the interaction of different descriptors which were involved in the generation of best models. The interaction energies of descriptors were generated by means of methyl probe with positive charge 1. The values of interaction energy of different descriptors were considered for generation of bond and oppressed as descriptors to set proximity between set of molecules [9].

Conclusion

In the present work 3D QSAR model was developed by Vlife Sciences MDS QSAR Plus software and the developed model help to predict the features required for good anti malarial activity of quinoxaline. In this study 3D QSAR model developed by k-NN method associated with stepwise forward-backward selection method gives acceptable q^2 (0.7589) and pred_r^2 (0.4752) values as well the this model has satisfactory internal and external predictive power. The selected model reports that the electrostatic descriptor plays a significant role in determining biotic activity, Positive value in electrostatic descriptor represents the electron withdrawing groups that are favourable in that region and negative value represents the electron rich or electron donating groups that are advantageous in particular region. On the above mentioned model, the description for descriptor in particular region helps to design the new analogues of quinoxalines as antimalarial.

Acknowledgement

Authors are thankful to the Chhattisgarh Council of Science and Technology, Raipur Chhattisgarh for financial support under the mini research project grant programs respectively. Authors are also thankful to Faculty of Pharmaceutical Sciences Shri Shankaracharya Technical Campus Bhilai Chhattisgarh for providing necessary infrastructure of the conduction of the work.

References

1 WHO Expert Committee on Malaria, Technical Report Series, *Twentieth Report*, *World Health Organization*, Geneva (2000).
2 Hansch C, Kurup A, Garg R & Gao H, *Chem Rev*, 14 (2001) 619.
3 Gil A, Pabón A, Galiano S, Burguete A, Pérez-Silanes S, Deharo E, Monge A & Aldana I, *Molecules*, 19 (2014) 2166.
4 Tariq S, Somakala K & Amir M, *Eur J Med Chem*, 1 (2018) 542.
5 Vicente E, Lima L M, Bongard E, Charnaud S, Villar R, Solano B, Burguete A, Perez-Silanes S, Aldana I, Vivas L & Monge A, *Eur J Med Chem*, 1 (2008) 1903.
6 Molecular Design Suite, *Vlife Sciences Technologies Pvt. Ltd*, Pune, India (2004).
7 Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H & Tropsha A, *J Med Chem*, 20 (2002) 2811.
8 Vaishnav Y, Kashyap P & Deep Kaur C, *Curr Nanomed*, 7 (2017) 59.
9 Sanmati K J & Achal M, *J Theor Comput Sci*, 2 (2015) 1.