Integrating GUSAR and QSAR analyses for antimalarial activity of synthetic prodiginines against multi drug resistant strain

Devidas T. Mahajan, Vijay H. Masand, Komalsing N. Patil, Taibi Ben Hadda & Vesna Rastija

Medicinal Chemistry Research

ISSN 1054-2523
Volume 22
Number 5

Med Chem Res (2013) 22:2284-2292
DOI 10.1007/s00044-012-0223-7

Springer

Med Chem Res (2013) 22:2284–2292
DOI 10.1007/s00044-012-0223-7

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Integrating GUSAR and QSAR analyses for antimalarial activity of synthetic prodiginines against multi drug resistant strain

Devidas T. Mahajan · Vijay H. Masand · Komalsing N. Patil · Taibi Ben Hadda · Vesna Rastija

Received: 9 March 2012 / Accepted: 1 September 2012 / Published online: 22 September 2012

Abstract In the present study, we have carried out extensive GUSAR and conventional 3D QSAR analyses of 49 synthetic prodiginines possessing moderate to high activities against multi drug resistant strain of *Plasmodium falciparum*. 2D and 3D descriptors, various statistical parameters, viz. R^2, R^2_{adj}, standard error, Y-randomization, etc., were checked to build successful QSAR model. The best four parametric GA-MLR 3D-QSAR model was found to have $R^2_{train} = 0.84$; $R^2_{adj} = 0.83$. GUSAR analysis was performed to vindicate the QSAR results and get additional results. The consensus GUSAR model based on QNA descriptor is found to have $R^2_{train} = 0.80$ and $Q^2_{train} = 0.76$. The analyses reveal that certain groups/atoms like –F, benzylic $–CH_2–$ and $–OCH_3$ play crucial role in deciding the antimalarial activity of prodiginines. The analyses could be useful to improve the antimalarial activity of these biologically privileged molecules.

Keywords GUSAR · QSAR · Prodiginines · Antimalarial activity

Abbreviations

GUSAR	General unrestricted structure activity relationships
QSAR	Quantitative structure activity relationships
GA-MLR	Genetic algorithm multi linear regression
MDR	Multi drug resistant

D. T. Mahajan · V. H. Masand (✉) · K. N. Patil
Department of Chemistry, Vidya Bharati College, Camp, Amravati, Maharashtra, India
e-mail: vijaymasand@gmail.com; vijaymasand@rediffmail.com

T. B. Hadda
Laboratoire Chimie des Matériaux, FSO Université Mohammed Premier, 60000 Oujda, Morocco

V. Rastija
Department of Chemistry, Faculty of Agriculture, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

Introduction

Malaria, responsible for more than two million deaths every year (WHO, 2010), is becoming a global health challenge due to the emergence of multi drug resistant race of causative agent (Papireddy *et al.*, 2011). The necessity to modify the available marketed drugs or discover new therapeutics is essential to provide better medicines against malaria. In the last two decades, various chemicals, viz. xanthones, artimisinins, prodiginines, etc., have been synthetically modified and tested to discover new therapeutic agents against malaria (Masand *et al.*, 2010a, b; Mahajan *et al.*, 2010; Wang *et al.*, 2008; Pinto *et al.*, 2005).

Among the various promising agents, prodiginines (Fig. 1) have gained considerable attention due to moderate to high potency against *Plasmodium falciparum*, ease of synthesis, etc. (Papireddy *et al.*, 2011; Williamson *et al.*, 2006). Prodiginines, the oligopyrrole derivatives with two pyrrole rings directly coupled in tandem array, can be administered orally and result in effective parasite clearance. Despite of these advantages, the search for prodiginines with improved ADMET (absorption, distribution, metabolism, excretion and toxicity) profile persists (Papireddy *et al.*, 2011; Masand *et al.*, 2012a, b, c). Undoubtedly, the modern methods of drug designing, viz. QSAR and molecular docking could be useful in achieving these goals (Masand *et al.*, 2011; Jawarkar *et al.*, 2010; Mota *et al.*, 2009). QSAR, a chemometric technique for ligand-based drug designing, is useful in finding the structural features and pattern that sway biological activity

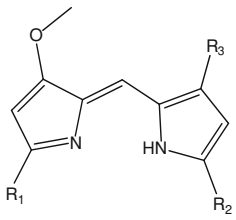


Fig. 1 Synthetic prodiginines used in present study

or a particular property. QSAR has been successfully applied for the prediction of biological activity before synthesis of untested compound, risk assessment, toxicity prediction, regulatory decisions, etc. (Masand *et al.,* 2010a, b, 2011, 2012a, b, c; Bohari *et al.*, 2011; Pasha *et al.*, 2007, Srivastava, 2008, Srivastava *et al.*, 2009). Molecular docking plays more useful role if the structure of receptor with which the drug interacts is known. Since the mechanism of action and the receptor with which prodiginines interact is unknown for antimalarial activity (Papireddy *et al.*, 2011; Masand *et al.*, 2012a, b, c), the QSAR and GUSAR analyses were performed to discover the structural features, which govern the antimalarial activity of prodiginines.

Materials and methods

Dataset and biological data

The 49 prodiginines assayed against *P. falciparum* pansensitive Dd2 (a multi drug resistant strain of *P. falciparum*) with chloroquine (CQ) as a reference drug for in vitro antimalarial activity were selected from literature (Papireddy *et al.*, 2011). The prodiginines selected for the study are structurally diverse. Antimalarial activities (nM, nanomolar) were converted into the corresponding $-\log_{10}$ IC_{50} values (pIC_{50}) to get symmetrically distributed data (Masand *et al.*, 2011), where IC_{50} value is the effective concentration of compound required to achieve 50 % of inhibition. The structures and in vitro activity of 49 prodiginines used in this study are listed in Table 1. For QSAR and GUSAR analysis, the dataset was divided into training and test set of 41 and 8 molecules, respectively.

Calculation, selection of descriptors and validation of QSAR model

ChemSketch software (ACD labs 12.0 freeware) was used to draw and optimize (MM+ method) 2D and 3D structures of the molecules using the default settings. The 3D optimized structures were used to calculate the descriptors. We tested different types of 3D descriptors like MoRSE, RDF, WHIM, GATEAWAY, etc., available in e-Dragon. For external validation of QSAR and GUSAR models, the dataset was divided into training and test set of 41 and 8 molecules, respectively. The optimum number and set of descriptors breed performance for a QSAR model (Masand *et al.*, 2010a, b) to achieve this goal Weka 3.6 was used. The thriving multi linear regression (MLR) equation based on four parameters was built by means of genetic algorithm (GA) following the standard procedure mentioned in the manual of Weka. Correlation matrix was constructed to avert the redundancy of descriptors (independent variables) and chance correlation (Masand *et al.*, 2010a, b). To prove that we have not arrived at 4-parametric equations inadvertently and the model is not a chance-comer as well, thorough statistical validation was performed. Correlation coefficient (R^2), R^2_{adj}, Q^2, F, s, etc., were calculated to bring the QSAR model in the arc of excellence (Masand *et al.*, 2010a, b; Jawarkar *et al.*, 2010).

Y-randomization test

Y-randomization is the most popular and probably the most powerful technique for the validation of a given QSAR model (Masand *et al.*, 2010a, b). In this approach, dependent variable vector (antimalarial activity in this study) is randomly shuffled and a new QSAR model is built using the original independent variables. The procedure is repeated number of times. If the new QSAR models have lower R^2 values for several trials, then the given QSAR model is thought to be robust. Thus, Y-randomization is useful to avoid any chance-comer correlation between dependent variable vector and independent variables. This Y-randomization was tested for model and low values of R^2 were observed (see Table 2).

GUSAR investigation

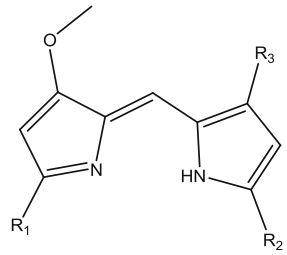
The GUSAR investigations were carried out by means of GUSAR 2011 by the standard procedure and settings (Kokurkina *et al.*, 2011; Masand *et al.*, 2011; Filimonov *et al.*, 1999, 2004, 2009; Poroikov *et al.*, 2004).

Results and discussions

QSAR analysis

QSAR is an established technique to discover the structural features which affect the biological profile of congener molecules (Masand *et al.*, 2010a, b, 2012a, b, c). An appropriately validated successful QSAR model can provide useful rational summary of structural features, which affect the biological activity. To be a successful QSAR model, it must be interpretable in terms of structural features (Masand *et al.*, 2010a, b). Therefore, careful

Table 1 Experimental data and predicted pIC_{50} by QSAR and GUSAR methods



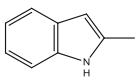
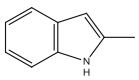
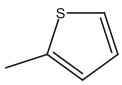
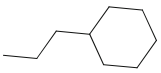
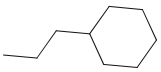
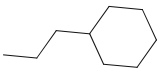
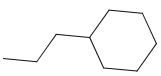
S. no.	R_1	R_2	R_3	IC_{50}(nM) Dd2	pIC_{50} expt.
1		CH_3	CH_3	4,950	−3.69461
2		n-$C_{11}H_{23}$	H	6,150	−3.78888
3		n-$C_{11}H_{23}$	H	15,750	−4.19728
4		CH_3	CH_3	13,640	−4.13481
5		n-$C_{11}H_{23}$	H	3,810	−3.58092
6		n-$C_{11}H_{23}$	H	7,770	−3.89042
7	2-Pyrolyl	n-C_4H_9	H	1,590	−3.20140
8	2-Pyrolyl	n-C_6H_{13}	H	450	−2.65321
9	2-Pyrolyl	n-C_8H_{17}	H	130	−2.11394
10	2-Pyrolyl	n-$C_{16}H_{33}$	H	400	−2.60206
11	2-Pyrolyl	H	$CH_2CH(CH_3)_2$	230	−2.36173
12	2-Pyrolyl	H	n-C_4H_9	18	−1.25527
13	2-Pyrolyl	H	n-C_6H_{13}	7	−0.84510
14	2-Pyrolyl	H	n-C_8H_{17}	1.8	−0.25527
15	2-Pyrolyl	H	n-$C_{10}H_{21}$	10	−1.00000
16	2-Pyrolyl	H	$C_6H_5CH_2$	86	−1.93450
17	2-Pyrolyl	H	4-$OCH_3C_6H_4CH_2$	156	−2.19312
18	2-Pyrolyl	H	4-$ClC_6H_4CH_2$	81	−1.90849
19	2-Pyrolyl	H	4-$BrC_6H_4CH_2$	108	−2.03342
20	2-Pyrolyl	CH_3	CH_3	8,130	−3.91009
21	2-Pyrolyl	n-C_6H_{13}	n-C_3H_7	4.0	−0.60206
22	2-Pyrolyl	n-C_8H_{17}	n-C_3H_7	2.7	−0.43136

Table 1 continued

S. no.	R_1	R_2	R_3	IC_{50}(nM) Dd2	pIC_{50} expt.
23	2-Pyrolyl	*n*-C_3H_7		1.3	−0.11394
24	2-Pyrolyl	*n*-C_6H_{13}	*n*-C_6H_{13}	1.1	−0.04139
25	2-Pyrolyl	*n*-C_7H_{15}	*n*-C_6H_{13}	1.2	−0.07918
26	2-Pyrolyl	*n*-C_6H_{13}	*n*-C_8H_{17}	2.0	−0.30103
27	2-Pyrolyl	*n*-C_7H_{15}	*n*-C_8H_{17}	2.9	−0.46240
28	2-Pyrolyl	*n*-C_8H_{17}	*n*-C_8H_{17}	129	−2.11059
29	2-Pyrolyl			3.5	−0.54407
30	2-Pyrolyl	C_2H_5	4-$ClC_6H_4CH_2$	6.2	−0.79239
31	2-Pyrolyl	*n*-C_3H_7	4-$ClC_6H_4CH_2$	2.6	−0.41497
32	2-Pyrolyl	*n*-C_6H_{13}	4-$ClC_6H_4CH_2$	1.8	−0.25527
33	2-Pyrolyl	*n*-C_7H_{15}	4-$ClC_6H_4CH_2$	2.2	−0.34242
34	2-Pyrolyl	*n*-C_8H_{17}	4-$ClC_6H_4CH_2$	12.0	−1.07918
35	2-Pyrolyl	4-$ClC_6H_4CH_2$		2.9	−0.46240
36	2-Pyrolyl	*n*-C_6H_{13}	4-$FC_6H_4CH_2$	0.9	0.045757
37	2-Pyrolyl	*n*-C_8H_{17}	4-$FC_6H_4CH_2$	1.2	−0.07918
38	2-Pyrolyl	*n*-C_6H_{13}	4-$BrC_6H_4CH_2$	2.8	−0.44716
39	2-Pyrolyl	*n*-C_8H_{17}	4-$BrC_6H_4CH_2$	2.9	−0.46240
40	2-Pyrolyl	4-$ClC_6H_4CH_2$	4-$ClC_6H_4CH_2$	4.8	−0.68124
41	2-Pyrolyl	4-$FC_6H_4CH_2$	4-$FC_6H_4CH_2$	5.7	−0.75587
42	2-Pyrolyl	4-$BrC_6H_4CH_2$	4-$BrC_6H_4CH_2$	11.0	−1.04139
43	2-Pyrolyl	4-$FC_6H_4CH_2$	4-$ClC_6H_4CH_2$	6.1	−0.78533
44	2-Pyrolyl	4-$BrC_6H_4CH_2$	4-$ClC_6H_4CH_2$	7.7	−0.88649
45	2-Pyrolyl	4-$BrC_6H_4CH_2$	4-$FC_6H_4CH_2$	5.1	−0.70757
46	2-Pyrolyl	2,4-$Cl_2C_6H_3CH_2$	2,4-$Cl_2C_6H_3CH_2$	11.0	−1.04139
47	2-Pyrolyl	2,4-$F_2C_6H_3CH_2$	2,4-$F_2C_6H_3CH_2$	18.3	−1.26245
48	2-Pyrolyl	3-$FC_6H_4CH_2$	3-$FC_6H_4CH_2$	6.7	−0.82607
49	2-Pyrolyl	2-$ClC_6H_4CH_2$	2-$ClC_6H_4CH_2$	4.9	−0.69020

determination, calculation and selection of descriptors are utmost important to build guiding QSAR model. The selected descriptors must be orthogonal; that is, they should have minimum correlation with each other (Jawarkar *et al.*, 2010). Correlation matrix was constructed to check correlation of the descriptors (Table 3).

We obtained improved model with better correlation statistics, when 2D and 3D descriptors were used in

Table 2 Value of R^2 after Y-randomization for QSAR model

Y-randomization	R^2 after Y-randomization for QSAR model
1	0.02
2	0.07
3	0.12
4	0.13
5	0.09

Table 3 Correlation matrix to check inter correlation of descriptors

Descriptor	*n*CbH	*F*02[C–O]	*F*06[C–O]	*F*07[SC–N]
*n*CbH	1.000			
*F*02[C–O]	0.101	1.000		
*F*06[C–O]	0.172	0.535	1.000	
*F*07*C–N*]	0.537	0.201	0.274	1.000

combination. The thriving four parameter-based GA-MLR equation along with various statistical characteristics is as follows:

$$pIC_{50} = -0.189(\pm 0.073)\, n\text{CbH} - 0.981(\pm 0.442)\, F02[\text{C-O}] + 1.445(\pm 0.600)\, F06[\text{C-O}] + 0.639(\pm 0.139)\, F07[\text{C-N}] - 9.310(\pm 2.288)$$

$$n_{train} = 41;\ R_{train} = 0.92;\ R^2_{train} = 0.84;\ R^2_{adj} = 0.83;\ s = 0.56;\ F_{train} = 48.13;\ p < 0.0001;\ Q^2_{train} = 0.80;\ \text{SPress} = 0.62;\ \text{SDEP} = 0.59;\ n_{test} = 8;\ R_{test} = 0.84;\ R^2_{test} = 0.71$$

In above model, the symbols have usual meaning (Masand *et al.*, 2010a, b; Jawarkar *et al.*, 2010). The high values of R, R^2, Q^2, F and low value of s indicate that model has excellent statistical significance. Moreover, the small difference between R^2_{adj} (which is considered to be a better parameter in judging the predictive power as compared to R^2) and R^2 confirms the high predictive power of model (Masand *et al.*, 2010a, b, 2011; Jawarkar *et al.*, 2010).

Interpretation of QSAR model

QSAR models are built to determine the structural features, which are responsible for the desired activity. *n*CbH corresponds to the number of unsubstituted benzene C (sp2). The negative coefficient for *n*CbH reveals that increase in number of unsubstituted aromatic C atoms is unfavourable for antimalarial activity of prodiginines and its value should be kept as low as possible. Another descriptor which affects negatively is *F*02[*C–O*] to biological activity. It is a 2D frequency fingerprint descriptor, which corresponds to the frequency of occurrence of C–O at a topological distance of two. It appears that *F*06[*C–O*] and *F*07[*C–N*] play positive roles in deciding the antimalarial activity. These are 2D frequency fingerprint descriptors, which correspond to frequency of occurrence of C–O and C–N at a topological distance of 6 and 7, respectively. This means that carbon should be present at a topological distance of 6 and 7 from oxygen and nitrogen, respectively. This finding gets firm support when we observe for compound number **36**, **24**, **37** and **25**. These results are further vindicated by GUSAR analysis.

GUSAR analysis

GUSAR is a relatively novel approach with considerable advantages over the conventional QSAR like the use of Self-Consistent Regression (SCR) algorithm for automatic selection of appropriate number and set of descriptors, unique type of descriptors that represent intermolecular interactions also (Kokurkina *et al.*, 2011; Masand *et al.*, 2011; Filimonov *et al.*, 1999, 2004, 2009; Poroikov *et al.*, 2004). GUSAR 2011 gives output in the form of a diagram containing various statistical characteristics for the consensus model, along with the atoms of each molecule coloured apropos of their contribution towards biological activity. The noticeable drawback of GUSAR 2011 is that it neither provides the QSAR model in interpretable form nor any knowledge about the descriptors that are used to build the consensus model.

The consensus GUSAR model was found to be with $n_{train} = 41, R^2_{train} = 0.80, F_{train} = 16.00, \text{SD} = 0.35, Q^2_{train} = 0.76, n_{test} = 8, R^2_{test} = 0.86$ and $V = 8$. The symbols have usual meaning (Kokurkina *et al.*, 2011; Masand *et al.*, 2011; Filimonov *et al.*, 1999, 2004, 2009; Poroikov *et al.*, 2004). If QSAR models were produced on the basis of QNA descriptors, the involvement of every atom into the predicted value is showed for a studied compound. The GUSAR analysis was carried out for all the 49 molecules. For the sake of representation only, the figure shows the GUSAR 2011 output for four most active and four least active compounds as representatives. The contribution is a calculation of activity value for a single atom from the structure of the studied molecule.

Plausible explanation of the colours is as following: "Green" means that the impact of the atom approximately corresponds to the predicted activity value for a whole molecule. "Blue" means that the particular atom may decrease the activity. "Red" means that the particular atom may increase the activity. Thus, if one would like to increase the activity, the number of "blue" atoms should be reduced, and the number of "red" atoms should be increased. One can analyse how many fragments have "red" and "blue" colours for finding the most important

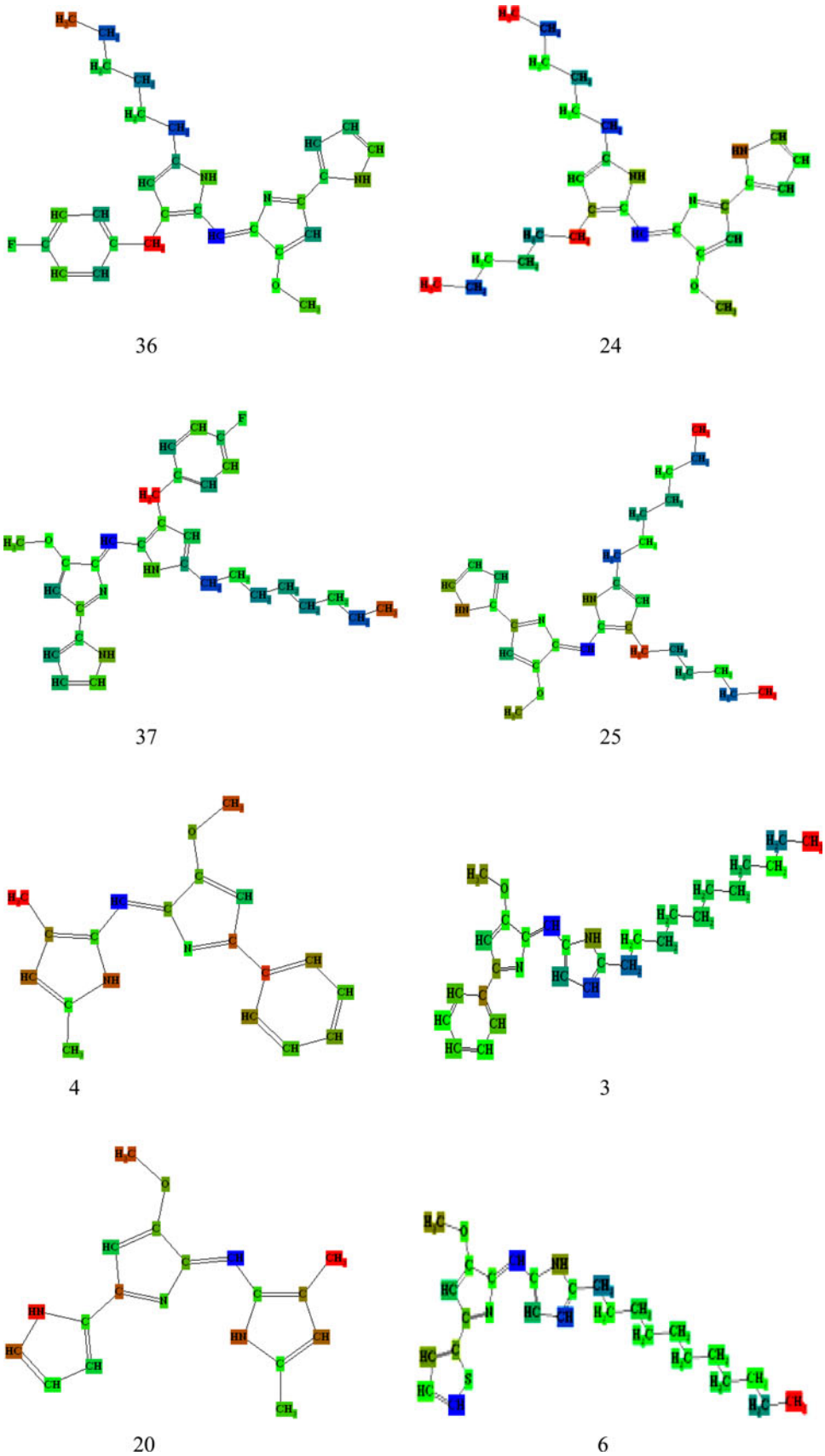


Fig. 2 GUSAR output for eight (four most active and four least active) compounds as representative (Color figure online)

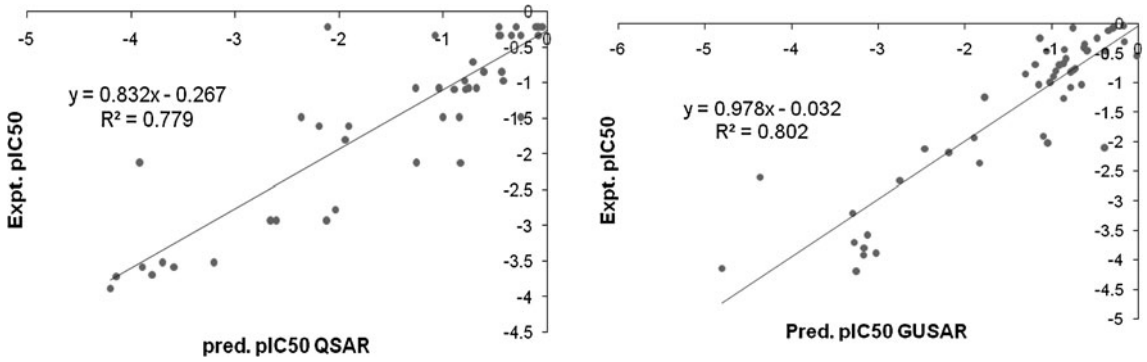


Fig. 3 Relation between experimental pIC_{50} and calculated pIC_{50} by QSAR and GUSAR models

Table 4 Descriptors used in QSAR model, pIC_{50} (experimental) and predicted pIC_{50} (By QSAR and GUSAR 2011)

S. no.	pIC_{50} obs	pIC_{50} QSAR model	Residual (expt. − pred. QSAR)	pIC_{50} GUSAR	Residual (expt. – pred. GUSAR)	*n*CbH	*F*02 [C–O]	*F*06 [C–O]	*F*07 [C–N]
1	−3.695	−3.525	−0.170	−3.2743	−0.4203	4	2	5	2
2	−3.789	−3.691	−0.098	−3.1556	−0.6333	4	2	4	4
3	−4.197	−3.880	−0.317	−3.2454	−0.9519	5	2	4	4
4	−4.135	−3.714	−0.421	−4.8036	0.6688	5	2	5	2
5	−3.581	−3.575	−0.006	−3.1269	−0.4540	0	5	6	3
6	−3.890	−3.575	−0.315	−3.0248	−0.8656	0	2	4	3
7	−3.201	−3.525	0.324	−3.2948	0.0934	0	2	4	3
8	−2.653[a]	−2.936	0.283	−2.7416	0.0884	0	2	4	4
9	−2.114[a]	−2.936	0.822	−2.4542	0.3403	0	2	4	4
10	−2.602	−2.936	0.334	−4.3636	1.7615	0	2	4	4
11	−2.362[a]	−1.491	−0.871	−1.8291	−0.5326	0	2	5	4
12	−1.255	−2.130	0.875	−1.7643	0.5090	0	2	5	3
13	−0.845	−1.491	0.646	−1.3022	0.4571	0	2	5	4
14	−0.255	−1.491	1.236	−1.1274	0.8721	0	2	5	4
15	−1.000	−1.491	0.491	−1.0113	0.0113	0	2	5	4
16	−1.935	−1.796	−0.139	−1.892	−0.0425	5	2	5	5
17	−2.193	−1.607	−0.586	−2.1769	−0.0162	4	4	6	5
18	−1.908	−1.607	−0.301	−1.0948	−0.8137	4	2	5	5
19	−2.033	−2.770	0.737	−1.0368	−0.9966	4	2	5	5
20	−3.910	−2.130	−1.780	−3.1707	−0.7394	0	2	5	2
21	−0.602[a]	−0.852	0.250	−0.8327	0.2307	0	2	5	5
22	−0.431	−0.852	0.421	−0.8553	0.4240	0	2	5	5
23	−0.114	−0.212	0.098	−0.3372	0.2233	0	2	5	6
24	−0.041	−0.212	0.171	−0.1694	0.1281	0	2	5	6
25	−0.079	−0.212	0.133	−0.2812	0.2020	0	2	5	6
26	−0.301	−0.212	−0.089	−0.1616	−0.1394	0	2	5	6
27	−0.462	−0.212	−0.250	−0.3797	−0.0826	0	2	5	6
28	−2.111	−0.212	−1.899	−0.3810	−1.7296	0	2	5	6
29	−0.544[a]	0.426	−0.970	−0.0166	−0.5275	0	2	5	7
30	−0.792	−0.968	0.176	−0.9503	0.1580	4	2	5	6
31	−0.415	−0.968	0.553	−0.6264	0.2115	4	2	5	6
32	−0.255	−0.328	0.073	−0.4774	0.2222	4	2	5	7
33	−0.342	−0.328	−0.014	−0.6199	0.2775	4	2	5	7
34	−1.079	−0.328	−0.751	−0.7837	−0.2955	4	2	5	7
35	−0.462	−0.328	−0.134	−0.5910	0.1287	4	2	5	7

Table 4 continued

S. no.	pIC_{50} obs	pIC_{50} QSAR model	Residual (expt. − pred. QSAR)	pIC_{50} GUSAR	Residual (expt. – pred. GUSAR)	*n*CbH	*F*02 [C–O]	*F*06 [C–O]	*F*07 [C–N]
36	0.046[a]	−0.33	0.376	−0.4728	0.5186	4	2	5	7
37	−0.079	−0.328	0.249	−0.7542	0.6751	4	2	5	7
38	−0.447	−0.328	−0.119	−0.5849	0.1378	4	2	5	7
39	−0.462	−0.328	−0.134	−1.0598	0.5974	4	2	5	7
40	−0.681	−1.084	0.403	−0.8632	0.182	8	2	5	7
41	−0.756	−1.084	0.328	−0.7297	−0.0261	8	2	5	7
42	−1.041	−1.084	0.043	−1.1477	0.1063	8	2	5	7
43	−0.785[a]	−1.086	0.301	−0.7409	−0.0444	8	2	5	7
44	−0.886[a]	−1.086	0.200	−0.9782	0.0917	8	2	5	7
45	−0.708	−0.706	−0.002	−0.9008	0.1932	8	2	5	7
46	−1.041	−1.084	0.043	−0.6547	−0.3867	6	2	5	7
47	−1.262	−1.084	−0.178	−0.8636	−0.3988	6	2	5	7
48	−0.826	−2.123	1.297	−0.7826	−0.0435	8	2	5	7
49	−0.690	−0.706	0.016	−1.1792	0.4890	8	2	5	7

[a] Test set

fragments (Kokurkina *et al.*, 2011; Masand *et al.*, 2011; Filimonov *et al.*, 1999, 2004, 2009; Poroikov *et al.*, 2004).

From Fig. 2, it is clear that the presence of –F at para position of aromatic ring (at R_3 position) has insignificant impact on activity. The –CH= , which is a bridge between two pyrrole rings, has negative influence on activity. The benzylic $–CH_2–$ group (at position R_3) plays positive role in deciding the activity. $–OCH_3$ group (present on middle pyrrole ring) has negative effect on activity. This is clear that these types of results are very difficult to obtain by means of conventional QSAR, but GUSAR is helpful in finding such results.

Figure 3 shows the relation between experimental pIC_{50} and calculated pIC_{50} by QSAR and GUSAR models.

The pIC_{50} predicted by GUSAR and QSAR models along with the values of four descriptors are listed in Table 4.

Conclusions

From the present study, it is evident that antimalarial activity of prodiginines varies with the frequency of occurrence of carbon from oxygen and nitrogen at topological distance of 6 and 7, respectively. Moreover, substitution at aromatic C atoms is favourable for activity. The presence of –F at para position of aromatic ring (attached to R_3 position) has negative impact on activity. The benzylic $–CH_2–$ group (position R_3) play positive role in deciding the activity.

Acknowledgments Authors are thankful to e-Dragon, ACD ChemSketch 12 freeware (www.acdlabs.com), Weka 3.7 and GUSAR 2010 developing teams for providing free or trial versions of their softwares. Sincere thanks are due to Dr. F. C. Raghuwanshi, Principal, Vidya Bharati College, Amravati for his constant support and encouragement throughout the work.

Conflict of interest The authors declare that they have no conflict of interests.

References

Bohari MH, Srivastava HK, Sastry GN (2011) Analogue-based approaches in anti-cancer compound modelling: the relevance of QSAR models. Org Med Chem Lett 1(3):1–12

Filimonov DA, Poroikov VV, Borodina YV, Gloriozova TA (1999) Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. J Chem Inf Comput Sci 39:666–670

Filimonov DA, Akimov DV, Poroikov VV (2004) The method of Self-Consistent Regression for the quantitative analysis of relationships between structure and properties of chemicals. Pharm Chem J 38(1):21–24

Filimonov DA, Zakharov AV, Lagunin AA, Poroikov VV (2009) QNA based 'Star Track' QSAR approach SAR and QSAR. Environ Res 20(7–8):679–709

Jawarkar RD, Masand VH, Patil KN, Mahajan DT, Youssoufi MH, Hadda TB, Kumbhare SL (2010) 3D-QSAR study on coumarin analogues as potent inhibitors of MAO-B using a COMFA approach. Der Pharma Chem 2(6):302–310

Kokurkina GV, Dutov MD, Shevelev SA, Popkov SV, Zakharov AV, Poroikov VV (2011) Synthesis, antifungal activity and QSAR study of 2-arylhydroxynitroindoles. Eur J Med Chem 46: 4374–4382

Mahajan DT, Jawarkar RD, Patil KN, Masand VH, Nazerruddin GM (2010) 3D-QSAR studies on xanthone derivatives to understand pharmacological activities as MAO inhibitors. Der Pharma Chem 2(4):298–308

Masand VH, Jawarkar RD, Patil KN, Mahajan DT, Hadda TB, Kurhade GH (2010a) COMFA analysis and toxicity risk assessment of coumarin analogues as Mao-A inhibitors: attempting better insight in drug design. Der Pharm Lett 2(6):350–357

Masand VH, Jawarkar RD, Patil KN, Nazerruddin GM, Bajaj SO (2010b) Correlation potential of Wiener index vis-à-vis molecular refractivity, Antimalarial activity of xanthone derivatives. Org Chem Indian J 6(1):30–38

Masand VH, Jawarkar RD, Mahajan DT, Hadda TB, Manikrao AM, Khatale PN, Vyas JV (2011) Presuming the probable anti-inflammatory mechanism of ursolic acid: a plant derived pentacyclic triterpenoid, using molecular docking. J Comput Method Mol Des 1(2):9–13

Masand VH, Jawarkar RD, Mahajan DT, Hadda TB, Sheikh J, Patil KN (2012a) QSAR and CoMFA studies of biphenyl analogs of the anti-tuberculosis drug (6S)-2-nitro-6-{[4-(trifluoromethoxy) benzyl]oxy}-6,7-dihydro-5H-imidazo[2,1-b][1,3]oxazine(PA-824). Med Chem Res 21:2624–2629. doi:10.1007/s00044-011-9787-x

Masand VH, Mahajan DT, Patil KN, Chinchkhede KD, Jawarkar RD, Hadda TB, Alafeefy AA, Shibi IG (2012b) k-NN, quantum mechanical and field similarity based analysis of xanthone derivatives as α-glucosidase inhibitors. Med Chem Res 1–12. doi:10.1007/s00044-012-9995-z

Masand VH, Mahajan DT, Jawarkar RD, Patil KN, Hadda TB, Thakur SD, Rastija V (2012c) CoMSIA and POM analyses of anti-malarial activity of synthetic prodiginines. Bioorg Med Chem Lett 22:4827–4835

Mota SGR, Barros TFM, Castilho S (2009) 2D QSAR studies on a series of bifonazole derivatives with antifungal activity. J Braz Chem Soc 20(3):451–459

Papireddy K, Smilkstein M, Kelly JX, Shweta, Salem SM, Alhamadsheh M, Haynes SW, Challis GL, Reynolds KA (2011) Anti-malarial activity of natural and synthetic prodiginines. J Med Chem 54:5296–5306

Pasha FA, Srivastava HK, Srivastava A, Singh PP (2007) QSTR study of small organic molecules against *Tetrahymena pyriformis*. QSAR Combi Sci 26(1):69–84

Pinto MMM, Sousa ME, Nascimento MSJ (2005) Xanthone derivatives: new insights in biological activities. Curr Med Chem 12:2517–2538

Poroikov VV, Filimonov DA, Borodina YV, Gloriozova TA, Sitnikov VB, Sadovnikov SV, Sosnov AV (2004) Quantitative relationships between structure and delayed neurotoxicity of chemicals studied by the self-consistent regression method using the PASS program. Pharm Chem J 38(4):188–190

Srivastava HK (2008) A comparative QSPR study of alkanes with the help of computational chemistry. Bull Korean Chem Soc 29(1):67–76

Srivastava HK, Pasha FA, Mishra SK, Singh PP (2009) Novel applications of atomic softness and QSAR study of testosterone derivatives. Med Chem Res 18(6):455–466

Wang B, Liu Y, Ke Z, Cui J, Chen W, Ma L (2008) Synthesis inhibitory activities and QSAR study of xanthone derivatives α-glucosidase inhibitors. Bioorg Med Chem 16:7185–7192

WHO report, 2010

Williamson NR, Fineran PC, Leeper FJ, Salmond GPC (2006) The biosynthesis and regulation of bacterial prodiginines. Nat Rev 4:887–899