Current Research in Drug Discovery

Original Research Paper

Determination of Optimum Values of Descriptors to Set Filters for Synthetic Tri-Pyrrole Derivatives (Prodiginines) Against Multi Drug Resistant Strain of *Plasmodium Falciparum*

[1]Vijay H. Masand, [1]Devidas T. Mahajan, [2]Eslam Pourbasheer, [3]Taibi Ben Hadda, [4]Harsh Chauhan, [5]J.M. Gajbhiye and [6]A.M. Alafeefy

[1]*Department of Chemistry, Vidya Bharati College, Camp, Amravati, Maharashtra, India*
[2]*Department of Chemistry, Faculty of Science, University of Qom, Qom, Iran*
[3]*Laboratoire Chimie des Matériaux, Université Mohammed Premier, Oujda-60000, Morocco*
[4]*Creighton University, Omaha, NE, USA*
[5]*Division of Organic Chemistry, CSIR-National Chemical Laboratory, Pune 411008, India*
[6]*Department of Pharmaceutical Chemistry, Kind Saud University, Alkharj, Saudi Arabia*

Article history
Received: 06-11-2014
Revised: 15-11-2014
Accepted: 31-12-2014

Corresponding Author:
Vijay H. Masand
Department of Chemistry,
Vidya Bharati College, Camp,
Amravati, Maharashtra, India
Email: vijaymasand@gmail.com

Abstract: In the present study, we have carried out extensive non-linear Quantitative-Structure Activity Relationship (QSAR) analysis to correlate *in vitro* anti-malarial activity against multi drug resistant strain of *Plasmodium falciparum*. Forty-three synthetic prodiginines with different structural features were used for their potential antimalarial activity. Linear, bilinear, biexponential and parabolic equations were developed. These equations were compared to determine the optimum values of descriptors for very useful and easily interpretable descriptors. The optimum values of these descriptors could be helpful in finding and optimizing a good lead compound. Obtained correlations reveal that various factors like lipophilicity, molecular weight and number of bonds have non-linear relation with the anti-malarial activity.

Keywords: Prodiginines, Anti-Malarial Activity, Optimum/Desirability Values, Hybrid Inverse-QSAR

Introduction

Malaria, a dreadful vector-borne protozoal disease is responsible for more than two million deaths every year (WHO, 2012; http://www.who.int/malaria/en/). Developing a potent antimalarial compound is still a major challenge for the medicinal chemists (Biamonte *et al*., 2013). The situation is worsening with the rapid spread of multi drug resistant strains of causative agent (Biamonte *et al*., 2013; Mahajan *et al*., 2012; 2013; Mara *et al*., 2013; Masand *et al*., 2013b; Murugesan *et al*., 2013; Ojha and Roy, 2012; Papireddy *et al*., 2011). Therefore, there is essential need to curb this deadly disease either by modifying the existing marketed drugs or developing new therapeutic molecules. Different compounds like xanthones, artemisinins, prodiginines have been synthesized and tested to develop new potential remedies for malaria (Biamonte *et al*., 2013; Mahajan *et al*., 2012; 2013; Mara *et al*., 2013; Masand *et al*., 2013b; Murugesan *et al*., 2013; Ojha and Roy, 2012; Papireddy *et al*., 2011).

Prodiginines (Mahajan *et al*., 2012; 2013; Masand *et al*., 2013b; Papireddy *et al*., 2011), the oligopyrrole derivatives with a characteristic conjugated system, are promising anti-malarial agents (Fig. 1). These compounds have the ability to inhibit *Plasmodium falciparum* (*P. falciparum*) at very low concentrations. They show marked clearance of the protozoa parasite and can be effectively administered orally. Despite these advantages, search for a potent prodiginines with good Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) profile and improved ease of synthesis has resulted in limited success (Mahajan *et al*., 2012; 2013; Masand *et al*., 2013b; Papireddy *et al*., 2011). The progress can be expedited using the contemporary method of drug designing like QSAR, Molecular docking and Pharmacophore modelling. Of the above mentioned methods, QSAR is an established technique with good success in last few decades and is utilized in our research (Huang and Fan, 2011; Myint and Xie, 2010; Scior *et al*., 2009; Tropsha, 2010).

DOI: 10.3844/crddsp.2014.51.59

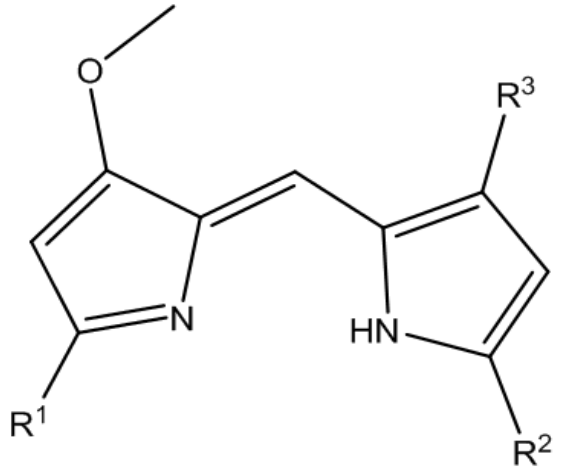

Fig. 1. Synthetic prodiginines used in present study

A typical QSAR study involves establishment of correlation between structure and activity (Mahajan *et al*., 2012; 2013; Masand *et al*., 2012a; Masand *et al*., 2013a; 2012b; 2013b; Rastija *et al*., 2013). Different characteristics or attributes of chemical structure are expressed in terms of numerical entities termed as molecular descriptors (also known as parameters or features). One or more molecular descriptors are used to build statistically robust linear regression equation. A properly validated QSAR equation is considered more useful if it is derived using descriptors that represent maximum useful information with minimum overlap and are interpretable in terms of structural features (Chirico and Gramatica, 2011; 2012; Chirico *et al*., 2012; Gramatica, 2013; Gramatica *et al*., 2012; 2013; Martin *et al*., 2012; Mitra *et al*., 2010; Roy and Mitra, 2012; Saha and Roy, 2012; Tropsha, 2010). Unfortunately, limited number of validated QSAR equations with the ability to guide for the development of new drugs or modification of existing drugs are utilized, due to the following reasons (Chirico and Gramatica, 2011; 2012; Chirico *et al*., 2012; Doweyko, 2008; Gramatica, 2013; Gramatica *et al*., 2012; 2013; Martin *et al*., 2012; Mitra *et al*., 2010; Roy and Mitra, 2012; Saha and Roy, 2012; Tropsha, 2010) (i) Difficulty in interpretation of QSAR equation in terms of structural features; (ii) The calculation or estimation of descriptors is very complex or resource consuming; (iii) computational facilities/resources like advanced and specific softwares may not be available to organic chemist to calculate descriptors that are mentioned in QSAR equation. (iv) The organic chemist may not be well skilled or trained in QSAR; (v) In addition, important descriptors having good correlation with activity might get missed in QSAR equations due to some reasons.

To overcome the difficulties, many researchers use inverse-QSAR (*i*-QSAR). In *i*-QSAR, the molecules are optimised using a set of physico-chemical properties or theoretical descriptors, which are obtained or derived using a well known marketed drug as 'reference' (Brown *et al*., 2006; Faulon *et al*., 2005). This approach has certain limitations like (i) proper selection of drug is an exigent and tricky process (ii) the drug should have similarity in structural or shape with the molecules of data set in hand (iii) For some diseases, no marketed drugs are available whereas for some diseases, a lot of marketed drugs are available. (iv) The physico-chemical properties or theoretical descriptors which are associated with one chemo-type of drug may not be possible to calculate or estimate for other chemo-type of molecule. (iv) The physico-chemical properties or theoretical descriptors associated with one chemo-type of drug may not be possible to calculate or estimate for other chemo-type molecules. After determining the values of different descriptors, the problem then lies in constructing a viable molecule from these descriptors. This is the real limiting factor of most inverse-QSAR methods, because most of the descriptors are not reversible.

A good solution is to determine the optimum value of useful and information rich descriptors, during the QSAR equation development. The most striking advantage in determining optimal values of different descriptors is that the 'most active' compound in the given data set may or may not fit to optimum values of all the descriptors. This optimization is not based on single 'reference' drug as in *i*-QSAR. In this case a data set is used to derive a set of physico-chemical properties or theoretical descriptors to optimize the molecules. Thereby, increasing the chances of finding better alternatives to visible 'most active' and potential compounds outside the present data set. This approach can be viewed as 'Hybrid-inverse QSAR'. It could significantly accelerate the discovery of novel small molecules with specified chemical properties.

Literature survey reveals (Buchwald and Yamashita, 2014; Gidskehaug *et al*., 2008; Hansch *et al*., 2004; Jager and Kooijman, 2009; Kubinyi, 2002) that a well established method to determine the optimum value of any descriptor is to derive non-linear equation, especially bilinear or biexponential or parabolic equation. These functions assume that the relationship between descriptor and the activity is non-linear with the vertex of curve representing the optimum value.

In our previous work, we successfully performed CoMSIA, GUSAR and QSAR analyses for antimalarial activity of synthetic prodiginines. The objectives of the present study are (i) to determine optimum value of easily interpretable descriptors used in the QSAR equation (ii) to determine optimum value of some other useful descriptors having good correlation with activity but not included in the reported QSAR equations and (iii) to compare the performance and ability of linear, parabolic, bilinear and biexponential QSAR models to determine the optimum values of descriptors.

DOI: 10.3844/crddsp.2014.51.59

Methodology

Data Set

The experimental *in vitro* Inhibitory Concentrations (IC_{50}) expressed in nanomolar units of forty three synthetic prodiginines against the Chloroquine (CQ) resistant strain Dd2 are selected from a recent publication (Papireddy *et al*., 2011). The data set includes prodiginines with different substituents like varying length of alkyl chain, substituents at different positions of benzene ring etc. Table 1, provides the experimental data. The values were converted into the logarithm units, ($-\log_{10} IC_{50} = pIC_{50}$) for molecular modelling purpose.

The structures were drawn using ACD Chemsketch 12 freeware and were converted into 3D structures. This was followed by geometry optimization using a molecular mechanics method implemented in the program VegaZZ, using Gasteiger partial charges and Tripos force field (Mahajan *et al*., 2012; 2013; Masand *et al*., 2013b). The optimized structures (β-isomer) were uploaded onto the e-DRAGON server to calculate myriad number of 1D-, 2D- and 3D- molecular descriptors (Fig. 2). Before QSAR model development, descriptors with constant or nearly constant (for 80% molecules) values were discarded. Genetic Algorithm (GA) available in QSARINS (Chirico and Gramatica, 2011; 2012; Chirico *et al*., 2012; 2013; Chirico and Gramatica, 2011; 2012; Chirico *et al*., 2012; 2013) was used to select optimum number and set of descriptors to build statistically sound multi linear regression equation. Matlab and BuildQSAR were used to build bilinear, biexponential and parabolic equations. In addition, Microsoft excel was used for different statistical functions.

A good number of statistical parameters like R, R^2, R^2_{adj}, S and F were calculated along with R^2_{cv} (R^2_{LOO}) for internal validation and to check the robustness of the model.

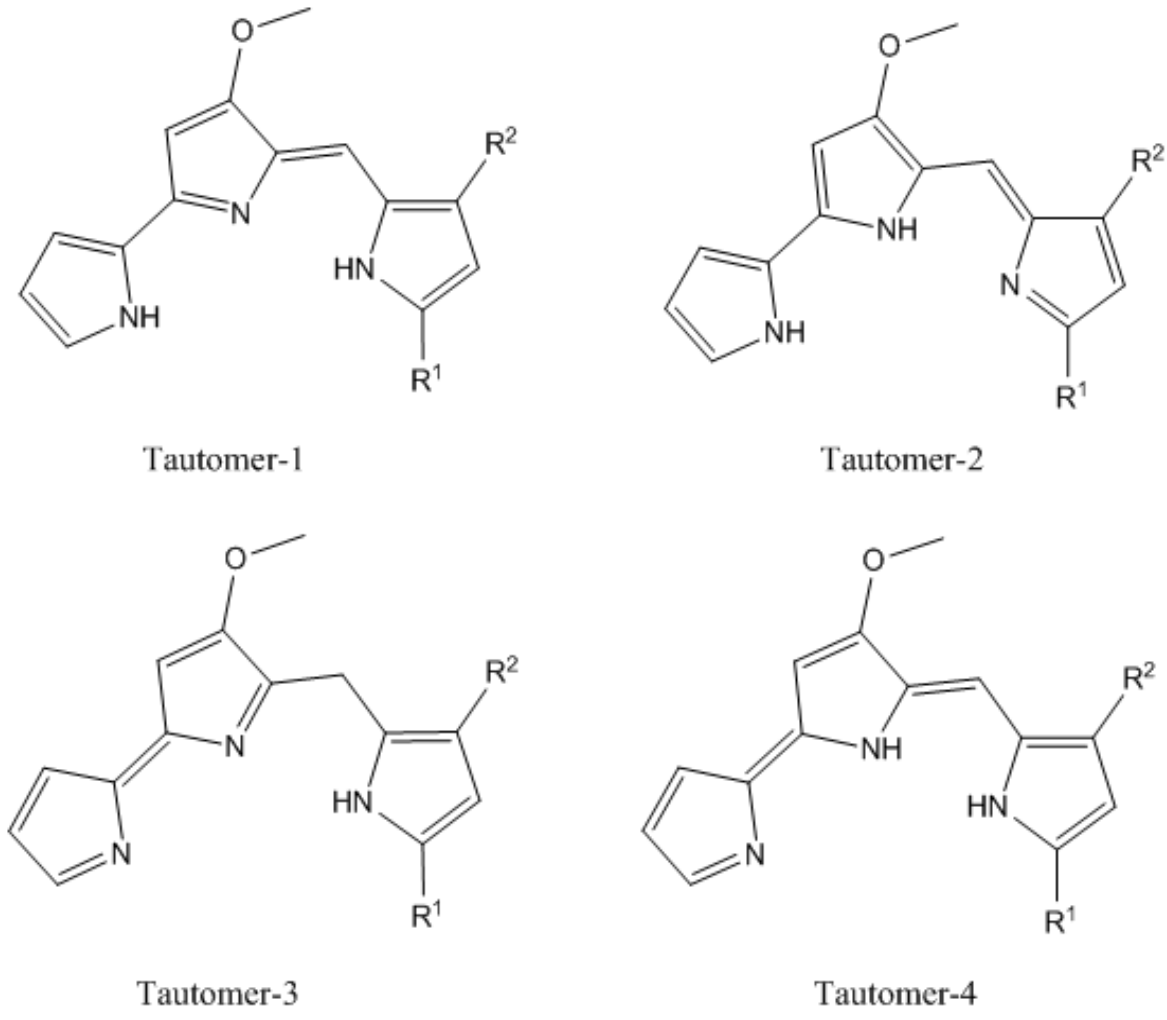

Fig. 2. Tautomeric forms of synthetic prodiginines (β-isomer) used in present study

DOI: 10.3844/crddsp.2014.51.59

Table 1. Different synthetic prodiginines along with experimental data IC_{50} (nM) and pIC_{50}

S. No.	R^1	R^2	R^3	IC_{50}(nM) Dd2	pIC_{50} expt.
1	2-pyrolyl	n-C_4H_9	H	1590.0	5.799
2	2-pyrolyl	n-C_6H_{13}	H	450.0	6.347
3	2-pyrolyl	n-C_8H_{17}	H	130.0	6.886
4	2-pyrolyl	n-$C_{16}H_{33}$	H	400.0	6.398
5	2-pyrolyl	H	$CH_2CH(CH_3)_2$	230.0	6.638
6	2-pyrolyl	H	n-C_4H_9	18.0	7.745
7	2-pyrolyl	H	n-C_6H_{13}	7.0	8.155
8	2-pyrolyl	H	n-C_8H_{17}	1.8	8.745
9	2-pyrolyl	H	n-$C_{10}H_{21}$	10.0	8.000[a]
10	2-pyrolyl	H	$C_6H_5CH_2$	86.0	7.066
11	2-pyrolyl	H	4-$OCH_3C_6H_4CH_2$	156.0	6.807
12	2-pyrolyl	H	4-$ClC_6H_4CH_2$	81.0	7.092
13	2-pyrolyl	H	4-$BrC_6H_4CH_2$	108.0	6.967
14	2-pyrolyl	CH_3	CH_3	8130.0	5.090
15	2-pyrolyl	n-C_6H_{13}	n-C_3H_7	4.0	8.398
16	2-pyrolyl	n-C_8H_{17}	n-C_3H_7	2.7	8.569
17	2-pyrolyl	n-C_3H_7		1.3	8.886
18	2-pyrolyl	n-C_6H_{13}	n-C_6H_{13}	1.1	8.959
19	2-pyrolyl	n-C_7H_{15}	n-C_6H_{13}	1.2	8.921
20	2-pyrolyl	n-C_6H_{13}	n-C_8H_{17}	2.0	8.699
21	2-pyrolyl	n-C_7H_{15}	n-C_8H_{17}	2.9	8.538
22	2-pyrolyl	n-C_8H_{17}	n-C_8H_{17}	129.0	6.889
23	2-pyrolyl			3.5	8.456
24	2-pyrolyl	C_2H_5	4-$ClC_6H_4CH_2$	6.2	8.208
25	2-pyrolyl	n-C_3H_7	4-$ClC_6H_4CH_2$	2.6	8.585
26	2-pyrolyl	n-C_6H_{13}	4-$ClC_6H_4CH_2$	1.8	8.745
27	2-pyrolyl	n-C_7H_{15}	4-$ClC_6H_4CH_2$	2.2	8.658
28	2-pyrolyl	n-C_8H_{17}	4-$ClC_6H_4CH_2$	12.0	7.921
29	2-pyrolyl	4-$ClC_6H_4CH_2$		2.9	8.538
30	2-pyrolyl	n-C_6H_{13}	4-$FC_6H_4CH_2$	0.9	9.046
31	2-pyrolyl	n-C_8H_{17}	4-$FC_6H_4CH_2$	1.2	8.921
32	2-pyrolyl	n-C_6H_{13}	4-$BrC_6H_4CH_2$	2.8	8.553
33	2-pyrolyl	n-C_8H_{17}	4-$BrC_6H_4CH_2$	2.9	8.538
34	2-pyrolyl	4-$ClC_6H_4CH_2$	4-$ClC_6H_4CH_2$	4.8	8.319
35	2-pyrolyl	4-$FC_6H_4CH_2$	4-$FC_6H_4CH_2$	5.7	8.244
36	2-pyrolyl	4-$BrC_6H_4CH_2$	4-$BrC_6H_4CH_2$	11.0	7.959
37	2-pyrolyl	4-$FC_6H_4CH_2$	4-$ClC_6H_4CH_2$	6.1	8.215
38	2-pyrolyl	4-$BrC_6H_4CH_2$	4-$ClC_6H_4CH_2$	7.7	8.114
39	2-pyrolyl	4-$BrC_6H_4CH_2$	4-$FC_6H_4CH_2$	5.1	8.292
40	2-pyrolyl	2,4-$Cl_2C_6H_3CH_2$	2,4-$Cl_2C_6H_3CH_2$	11.0	7.959
41	2-pyrolyl	2,4-$F_2C_6H_3CH_2$	2,4-$F_2C_6H_3CH_2$	18.3	7.738
42	2-pyrolyl	3-$FC_6H_4CH_2$	3-$FC_6H_4CH_2$	6.7	8.174
43	2-pyrolyl	2-$ClC_6H_4CH_2$	2-$ClC_6H_4CH_2$	4.9	8.310

Results and Discussion

In the present study, we derived and compared the linear, biexponential, bilinear (two equations) and parabolic equations. These equations are listed in Table 2 and 3. The equations provide useful correlation between activity and many easily interpretable useful descriptors like *Sv* (Sum of atomic van der Waal's volumes), *Sp* (Sum of atomic polarizabilities), *X1v* (first order valence connectivity index, to represent the steric factor), ALOG*P* (Ghose-Crippen Octanol-water coefficient) and *nAT* (number of atoms).

The linear model cannot be used for the determination of optimum value of any descriptor. The general form of the parabolic, bilinear Equation 1 (proposed by Kubinyi), bilinear Equation 2 and biexponential model is as following:

$$Y = aX + bX^2 + c \quad (1)$$

$$Y = aX + b \log(cX + 1) + d \quad (2)$$

$$Y = aX + b \, Log(\beta \, 10^X + 1) + c \quad (3)$$

$$Y = -b \, (\log((e^{(-c(x-a)/b)})+(e^{(e(x-a)/b))})) + d \quad (4)$$

Comparison of Different Models

For some descriptors viz. *nAT*, *Sv*, *Sp* and *ALOGP*, non-linear models are either superior or equivalent to the linear model. Whereas, for rest of the descriptors, the fitting of the non-linear models is better than the linear model. This indicates that the relation between the activity and the selected descriptors is non-linear in nature. In other words, non-linear model can better explain the variation of activity. Among non-linear models, bilinear Equation 1 (based on Kubinyi formula) fits better than the rest, with biexponential models being least fit in nature for many descriptors. None of the model satisfies the recommended threshold value (>0.85) of *CCC*, though for some models, it is close to it.

In the above models, the symbols have their usual meanings. Increasing the number of congeneric compounds in the data set as well as the range of biological data might result in better statistical fitting.

In many cases, the substantial fitting of the equation (R^2>0.60), though not outstanding, is satisfactory. This proves that there prevails the optimum value of lipophilicity, number of atoms, number of bonds and *X1v* (to represent steric factor). Thus, the selected descriptors, for which the optimum values are determined, represent the overall descriptor space.

A comparison of values of descriptors for four most active (highlighted as bold and italic) and four least active (highlighted as bold and italic) compounds justify the importance of optimum values of descriptors (Table 3). The value for selected descriptors for four most active molecules selected as representatives are close to optimum values whereas reverse is true for the four least active molecules. Thus, the optimum values of these descriptors could be helpful in finding a good "lead prodiginine" for anti-malarial activity.

Interestingly, the values of descriptors for the 'most active' compound 30 in the present data set are close to optimum values of many descriptors. However, it does not match with the optimum values of all the descriptors. This confirms that the appropriate lead/drug optimization using only most active or single drug as 'reference' is not a perfect method.

Table 2. Different linear and non-linear equations along with their statistical parameters

Descriptor	Statistical parameter	Linear	Parabolic	Bilinear Equation 1	Bilinear Equation 2	Biexponential
nAT	R^2	0.242	0.685	0.640	0.664	0.242
	R^2_{adj}	0.224	0.669	0.612	0.638	0.162
	RMSE	0.800	0.516	0.565	0.533	0.800
	SSE	27.517	11.442	13.073	12.211	27.517
	F	13.099	43.465	23.105	25.349	3.035
	CCC	0.389	0.813	0.781	0.796	0.390
nBT	R^2	0.267	0.676	0.657	0.654	0.713
	R^2_{adj}	0.249	0.660	0.631	0.628	0.683
	RMSE	0.786	0.523	0.565	0.540	0.492
	SSE	26.613	11.758	12.443	12.548	10.407
	F	14.935	41.756	24.933	24.354	23.629
	CCC	0.421	0.807	0.793	0.790	0.833
Sv	R^2	0.331	0.633	0.673	0.612	0.331
	R^2_{adj}	0.314	0.614	0.648	0.582	0.260
	RMSE	0.752	0.557	0.551	0.573	0.752
	SSE	24.302	13.339	11.857	14.093	24.302
	F	20.254	34.438	26.808	20.279	4.693
	CCC	0.497	0.775	0.805	0.758	0.497
Sp	R^2	0.320	0.647	0.679	0.624	0.320
	R^2_{adj}	0.304	0.629	0.654	0.596	0.249
	RMSE	0.758	0.546	0.547	0.563	0.758
	SSE	24.679	12.819	11.652	13.639	24.679
	F	19.317	36.648	27.508	21.377	4.475
	CCC	0.485	0.786	0.809	0.767	0.485
X1v	R^2	0.329	0.625	0.641	0.605	NC
	R^2_{adj}	0.313	0.607	0.614	0.574	NC
	RMSE	0.752	0.562	0.578	0.578	NC
	SSE	24.336	13.601	13.022	14.351	NC
	F	20.169	33.388	23.245	19.717	NC
	CCC	0.496	0.769	0.782	0.753	NC
ALOGP	R^2	0.323	0.683	0.711	0.650	0.232
	R^2_{adj}	0.307	0.667	0.688	0.623	0.151
	RMSE	0.756	0.517	0.519	0.543	0.805
	SSE	24.573	11.506	10.511	12.701	27.875
	F	19.579	43.112	31.906	23.576	9.349
	CCC	0.488	0.812	0.831	0.786	0.065

SSE- Sum of Squared Errors, RMSE-Root Mean Square Error, CCC- Concordance Correlation Coefficient

In present case, a plausible reason for this could be the ability of the molecules (prodiginines in present case) to attain different conformations and tautomeric forms. Prodiginines possess azafulvene-pyrrole tautomerism due to the three pyrrole rings joined by -CH= link. As prodiginines can form four different tautomeric forms, the tautomeric form, which is energetically favoured in solution, may not be the 'bioactive tautomeric form' which shows interaction with the specific receptor and is responsible for the pharmacologic activity of this group.

Prodiginine may interact with different receptors in different tautomeric forms. In addition, prodiginines can exist in two conformations viz. α and β isomer, which have been discussed in our previous work (Reference). Another possible reason is satisfactory fitting (R^2~0.60) for most of the developed models.

Variation of Activity with Various Parameters

Herein, the activities of some more active and less active molecules from the dataset in terms of various descriptors like lipophilicity/hydrophobicity, number of rotatable bonds, steric factor etc. for which the optimum value, determined using bilinear Equation 2, has been derived and discussed. For optimum value determination, parabolic and bilinear Equation 1 can also be use, but, these have some serious drawbacks, like (1) the parabolic approach forces the data into a symmetrical parabola, resulting in deviations between the experimental and parabola-calculated data. (2) The ascending slope is curved and conflicts with the observed linear data. (3) The bilinear equation provides better optimum value only if the dataset is large in size with wide spread variation in activity value. The bilinear Equation 2 does not confined to such limitations. Therefore, in the present work, it has been used for optimum value determination.

We here clarify that we have though discussed the effect of individual descriptor, but the combined or converse effect of other factors/descriptors do have additional influence on the activity profile of these compounds.

nBT (Number of Bonds)

The optimum value for the number of bonds from the bilinear equation (Table 2) is 66.076. This suggests that the compounds that have number of bonds closer to this value should have good activity compared tothe rest of the compounds. This observation is supported by the lower activity of the following compounds which possess either very low or very high *nBT*: 1 (*nBT* = 45, IC_{50} = 1590 nM), 2 (*nBT* = 51, IC_{50} = 450 nM), 4 (*nBT* = 81, IC_{50} = 400 nM), 11 (*nBT* = 51, IC_{50} = 156 nM), 14 (*nBT* = 39, IC_{50} = 8130 nM) and 22 (*nBT* = 81, IC_{50} = 129 nM). A comparison of following pairs of compounds further confirms this observation: 15 (*nBT* = 60, IC_{50} = 4.0 nM) with 16 (*nBT* = 66, IC_{50} = 2.7 nM), 6 (*nBT* = 45, IC_{50} = 18 nM) with 7 (*nBT* = 51, IC_{50} = 7 nM) with 8 (*nBT* = 57, IC_{50} = 1.8 nM). Though, compound number 10 possess *nBT* = 63 (close to optimum value) but its activity is very low with IC_{50} = 129 nM. This could be attributed to high value of *F10[C-C]*, which has negative contribution towards the activity profile. Another examples are 20 (*nBT* = 75, IC_{50} = 2.0 nM), 21 (*nBT* = 78, IC_{50} = 2.9 nM) and 22 (*nBT* = 81, IC_{50} = 129 nM). In addition, similar trend is observed for 24 (*nBT* = 53, IC_{50} = 6.2 nM), 25 (*nBT* = 56, IC_{50} = 2.6 nM), 26 (*nBT* = 65, IC_{50} = 1.8 nM), 27 (*nBT* = 68, IC_{50} = 2.2 nM) and 28 (*nBT* = 71, IC_{50} = 12.0 nM). The most active compound 30 (IC_{50} = 0.9 nM) is with *nBT*= 65, which is very close to optimum value.

Sv (Sum of Atomic Van Der Waal's Volumes)

The optimum value for *Sv* is 41.753. The two most active and two least active compounds 30 (IC_{50} = 0.9 nM), 31(IC_{50} = 1.2 nM), 1(IC_{50} = 1590 nM) and 2 (IC_{50} = 450 nM) have *Sv* = 38.97, 42.17, 26.87 and 30.07, respectively. For the active compounds the value of *Sv* is close to the optimum value, while reverse is true for the least active molecules. This observation is further supported by low activity of 14 (*Sv* = 23.68, IC_{50} = 8130 nM), 4 (*Sv* = 46.05, IC_{50} = 400 nM), 22 (*Sv* = 46.05, IC_{50} = 129 nM) and 5 (*Sv* = 26.87, IC_{50} = 230 nM).

Sp (Sum of Atomic Polarizabilities)

For this descriptor *Sp*, the optimum value obtained from the bi-linear equation is 44.281. The most active compounds 30 (IC_{50} = 0.9 nM), 31(IC_{50} = 1.2 nM) and the least active compounds 1(IC_{50} = 1590 nM) and 2 (IC_{50} = 450 nM) have *Sp* = 41.07, 44.59, 28.32 and 31.85, respectively. In addition, compounds4 and 14 have *Sp* = 49.46 and 24.80 with IC_{50} = 400 and 8130 nM, respectively.

X1v (First Order Valence Connectivity Index, to Represent the Steric Factor)

For *X1v*, the optimum value obtained from the bi-linear equation is 11.790. The most active compounds 30 (IC_{50} = 0.9 nM), 31(IC_{50} = 1.2 nM) and the least active compounds 1(IC_{50} = 1590 nM) and 2 (IC_{50} = 450 nM) have *X1v* = 11.314, 12.314, 7.680 and 8.680, respectively. In addition, compound 4 and 14 possess *X1v* = 13.680 and 6.536 with IC_{50} = 400 and 8130 nM, respectively.

Lipophilicity/Hydrophobicity (in Terms of ALOGP)

In modern drug designing, lipophilicity is considered as one of the most important factors. For the present data set, the optimum value of ALOGP is 7.112 from a parabolic equation. Similar to other descriptors, the most active compounds 30 (IC_{50} = 0.9 nM), 31(IC_{50} = 1.2 nM) and the least active compounds 1(IC_{50} = 1590 nM) and 2

(IC_{50} = 450 nM) have *ALOGP* = 6.970, 7.882, 3.878 and 4.790, respectively. In addition, compound 4 and 14 have *ALOGP* = 9.352 and 2.785 with IC_{50} = 400 and 8130 nM, respectively. This means, the compounds that posses number of bonds closer to this value should have good activity than the rest of the compounds. This observation is supported by the lower activity of following compounds which possess either very low or very high *ALOGP*: 11 (*ALOGP* = 4.111, IC_{50} = 156 nM) and 22 (*ALOGP* = 9.382, IC_{50} = 129 nM). A comparison of following pairs of compounds further confirms this observation: 15 (*ALOGP* = 6.189, IC_{50} = 4.0 nM) with 16 (*ALOGP* = 7.101, IC_{50} = 2.7 nM), 6 (*ALOGP* = 4.008, IC_{50} = 18 nM) with 7 (*ALOGP* = 4.920, IC_{50} = 7 nM) with 8 (*ALOGP* = 5.832, IC_{50} = 1.8 nM). Another example is 3 (*ALOGP* = 5.702, IC_{50} = 130 nM), 16 (*ALOGP* = 7.101, IC_{50} = 2.7 nM) and 22 (*ALOGP* = 9.382, IC_{50} = 129 nM). In addition, similar trend is observed for 24 (*ALOGP* = 5.604, IC_{50} = 6.2 nM), 25 (*ALOGP* = 6.060, IC_{50} = 2.6 nM), 26 (*ALOGP* = 7.429, IC_{50} = 1.8 nM), 27 (*ALOGP* = 7.885, IC_{50} = 2.2 nM) and 28 (*ALOGP* = 8.341, IC_{50} = 12.0 nM).

nAT (Number of Atoms)

Similar to *nBT*, this is a very easily interpretable and a useful descriptor for synthetic chemists. From parabolic equation, the optimum value obtained is 63.031. The active molecules possess *nAT* close to the optimum value, whereas opposite is true for the less active molecules. Examples are 1 (*nAT* = 43, IC_{50} = 1590 nM), 2 (*nAT* = 49, IC_{50} = 450 nM), 4 (*nAT* = 79, IC_{50} = 400 nM), 11 (*nAT* = 48, IC_{50} = 156 nM), 14 (*nAT* = 37, IC_{50} = 8130 nM) and 22 (*nAT* = 79, IC_{50} = 129 nM). A comparison of following pairs of compounds further confirms this observation: 15 (*nAT* = 58, IC_{50} = 4.0 nM) with 16 (*nAT* = 64, IC_{50} = 2.7 nM), 6 (*nAT* = 44, IC_{50} = 18 nM) with 7 (*nAT* = 49, IC_{50} = 7 nM) with 8 (*nAT* = 55, IC_{50} = 1.8 nM). Other examples are 20 (*nAT* = 73, IC_{50} = 2.0 nM), 21 (*nAT* = 76, IC_{50} = 2.9 nM) and 22 (*nAT* = 79, IC_{50} = 129 nM). In addition, similar trend is observed for 24 (*nAT* = 50, IC_{50} = 6.2 nM), 25 (*nAT* = 53, IC_{50} = 2.6 nM), 26 (*nAT* = 62, IC_{50} = 1.8 nM), 27 (*nAT* = 65, IC_{50} = 2.2 nM) and 28 (*nAT* = 68, IC_{50} = 12.0 nM). To add further, the most active compound 30 (IC_{50} = 0.9 nM) is with *nAT*= 62, which is very close to the optimum value.

Conclusion

In summary, the present study reveals that the non-linear models should be developed to determine optimum values of the descriptors. A good lead compound (prodiginine in the present work) can be identified and optimized if the optimum value of lipophilicity, sum of atomic van der Waal's volumes, sum of atomic polarizabilities, first order valence connectivity index, number of atoms, number of benzene-like rings and number of rotatable bond are used correctly and efficiently. The "ready to use" optimum/desirability values will be useful to the medicinal chemists in developing novel prodiginines with good anti-malarial activity profile.

Acknowledgement

We are thankful to e-Dragon, Vega ZZ, ACD ChemSketch, BILIN and RapidMiner developing teams for providing free/trial versions of their softwares. VHM is thankful to Dr. Paola Gramatica and QSARINS developing team for providing free copy of QSARINS.

Author's Contributions

Vijay H. Masand: Performed QSAR analysis and designed the research plan and organized the study.

Devidas T. Mahajan: Performed QSAR analysis and manuscript writing.

Eslam Pourbasheer: Performed QSAR analysis and manuscript writing.

Taibi Ben Hadda: Structure drawing and contributed in writing the results and discussion section.

Harsh Chauhan: Calculation of Descriptors and revised English of the manuscript.

J.M. Gajbhiye: Contributed in writing the manuscript.

A.M. Alafeefy: Contributed in writing the results and discussion section and designed the research plan and organized the study.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Biamonte, M.A., J. Wanner, K.G. Le Roch, 2013. Recent advances in malaria drug discovery. Bioorg. Med. Chem. Lett., 23: 2829-2843. DOI: 10.1016/j.bmcl.2013.03.067

Brown, N., B. McKay and J. Gasteiger, 2006. A novel workflow for the inverse QSPR problem using multiobjective optimization. J. Comp. Aided Mol. Des., 20: 333-341. DOI: 10.1007/s10822-006-9063-1

Buchwald, P. and F. Yamashita, 2014. Bilinear model for the size-dependency of the CYP3A4 inhibitory activity of structurally diverse compounds. Mol. Inf., 33: 8-14. DOI: 10.1002/minf.201300132

Chirico, N. and P. Gramatica, 2011. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. J. Chem. Inf. Model., 51: 2320-2335.

Chirico, N. and P. Gramatica, 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J. Chem. Inf. Model., 52: 2044-2058.

Chirico, N., E. Papa, S. Kovarich, S. Cassani and P. Gramatica, 2012. QSARINS, software for QSAR MLR model development and validation. University of Insubria, Varese, Italy.

Doweyko, A.M., 2008. QSAR: Dead or alive? J. Comput. Aided Mol. Des., 22: 81-89. DOI: 10.1007/s10822-007-9162-7

Faulon, J.L., W. Brown and S. Martin, 2005. Reverse engineering chemical structures from molecular descriptors: How many solutions? J. Comp. Aided Mol. Des., 19: 637-650. DOI: 10.1007/s10822-005-9007-1

Gidskehaug, L., E. Anderssen and B.K. Alsberg, 2008. Cross model validation and optimisation of bilinear regression models. Chemo. Intell. Lab. Syst., 93: 1-10. DOI: 10.1016/j.chemolab.2008.01.005

Gramatica, P., 2013. On the development and validation of QSAR models. Methods Mol. Biol., 930: 499-526. DOI: 10.1007/978-1-62703-059-5_21

Gramatica, P., N. Chirico, E. Papa, S. Cassani and S. Kovarich, 2013. QSARINS: A new software for the development, analysis and validation of QSAR MLR models. J. Comput. Chem., 34: 2121-2132. DOI: 10.1002/jcc.23361

Gramatica, P., S. Cassani, P.P. Roy, S. Kovarich and C.W. Yap *et al*., 2012. QSAR modeling is not "Push a Button and Find a Correlation": A case study of toxicity of (Benzo-)triazoles on Algae. Mol. Inf., 31: 817-835. DOI: 10.1002/minf.201200075

Hansch, C., A. Leo, S.B. Mekapati and A. Kurup, 2004. QSAR and ADME. Bioorg. Med. Chem. Lett., 12: 3391-3400. DOI: 10.1016/j.bmc.2003.11.037

Huang, J. and X. Fan, 2011. Why QSAR fails: An empirical evaluation using conventional computational approach. Mol. Pharm., 8: 600-608. DOI: 10.1021/mp100423u

Jager, T. and S.A.L.M. Kooijman, 2009. A biology-based approach for Quantitative Structure-Activity Relationships (QSARs) in ecotoxicity. Ecotoxicology, 18: 187-196. DOI: 10.1007/s10646-008-0271-4

Kubinyi, H., 2002. From narcosis to hyperspace: The history of QSAR. Quant. Struct. Act. Rel., 21: 348-356. DOI: 10.1002/1521-3838(200210)21:4<348::AID-QSAR348>3.0.CO;2-D

Mahajan, D.T., V.H. Masand, K.N. Patil, T. Ben Hadda and R.D. Jawarkar *et al*., 2012. CoMSIA and POM analyses of anti-malarial activity of synthetic prodiginines. Bioorg. Med. Chem. Lett., 22: 4827-4835. DOI: 10.1016/j.bmcl.2012.05.115

Mahajan, D.T., V.H. Masand, K.N. Patil, T.B. Hadda and V. Rastija, 2013. Integrating GUSAR and QSAR analyses for antimalarial activity of synthetic prodiginines against multi drug resistant strain. Med. Chem. Res., 22: 2284-2292. DOI: 10.1007/s00044-012-0223-7

Mara, C., E. Dempsey, A. Bell and J.W. Barlow, 2013. Synthesis and evaluation of phenoxyoxazaphospholidine, phenoxyoxazaphosphinane and benzodioxaphosphininamine sulfides and related compounds as potential anti-malarial agents. Bioorg. Med. Chem. Lett., 23: 3580-3583. DOI: 10.1016/j.bmcl.2013.04.026

Martin, T.M., P. Harten, D.M. Young, E.N. Muratov and A. Golbraikh *et al*., 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model., 52: 2570-2578.

Masand, V.H., R.D. Jawarkar, D.T. Mahajan, T.B. Hadda and J. Sheikh *et al*., 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, V.H., D.T. Mahajan, K.N. Patil, K.D. Chinchkhede and R.D. Jawarkar *et al*., 2012b. k-NN, quantum mechanical and field similarity based analysis of xanthone derivatives as α-glucosidase inhibitors. Med. Chem. Res., 21: 4523-4534. DOI: 10.1007/s00044-012-9995-z

Masand, V.H., D.T. Mahajan, T. Ben Hadda, R.D. Jawarkar and H. Chavan, 2013a. Molecular Docking and Quantitative Structure Activity Relationship (QSAR) analyses of indolylarylsulfones as HIV-1 non-nucleoside reverse transcriptase inhibitors. Med. Chem. Res. DOI: 10.1007/s00044-013-0647-8

Masand, V.H., D.T. Mahajan, K.N. Patil, T.B. Hadda and M.H. Youssoufi *et al*., 2013b. Optimization of antimalarial activity of synthetic prodiginines: QSAR, GUSAR and CoMFA analyses. Chem. Biol. Drug. Des., 81: 527-536. DOI: 10.1111/cbdd.12099

Mitra, I., P.P. Roy, S. Kar, P.K. Ojha and K. Roy, 2010. On further application of r_m^2 as a metric for validation of QSAR models. J. Chemomem., 24: 22-33. DOI: 10.1002/cem.1268

Murugesan, D., A. Mital, M. Kaiser, D.M. Shackleford and J. Morizzi *et al*., 2013. Discovery and structure-activity relationships of pyrrolone antimalarials. J. Med. Chem., 56: 2975-2990. DOI: 10.1021/jm400009c

Myint, K.Z. and X.Q. Xie, 2010. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. Int. J. Mol. Sci., 11: 3846-3866. DOI: 10.3390/ijms11103846

Ojha, P.K. and K. Roy, 2012. First report on exploring structural requirements of 1,2,3,4-Tetrahydroacridin-9(10H)-one analogs as antimalarials using multiple QSAR approaches: Descriptor-based QSAR, CoMFA-CoMSIA 3DQSAR, HQSAR and G-QSAR approaches. Comb. Chem. High Throughput Screen, 16: 7-21.

Papireddy, K., M. Smilkstein, J.X. Kelly S.M. Salem and M. Alhamadsheh *et al*., 2011. Antimalarial activity of natural and synthetic prodiginines. J. Med. Chem., 54: 5296-5306. DOI: 10.1021/jm200543y

Rastija, V., S. Nikolic and V.H. Masand, 2013. Quantitative relationships between structure and lipophilicity of naturally occurring polyphenols. Acta Chim. Slov., 60: 781-789.

Roy, K. and I. Mitra, 2012. On the use of the metric rm 2 as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. Mini. Rev. Med. Chem., 12: 491-504. DOI: 10.2174/138955712800493861

Saha, A. and K. Roy, 2012. In silico modeling for prediction of drug-induced adverse reactions and environmental hazards using QSAR tools. Curr. Drug Saf., 7: 255-256. DOI: 10.2174/157488612804096560

Scior, T., J.L. Medina-Franco, Q.T. Do, K. Martinez-Mayorga and J.A. Yunes Rojas *et al*., 2009. How to recognize and workaround pitfalls in QSAR studies: A critical review. Curr. Med. Chem., 16: 4297-4313. DOI: 10.2174/092986709789578213

Tropsha, A., 2010. Best practices for QSAR model development, validation and exploitation. Mol. Inform., 29: 476-488. DOI: 10.1002/minf.201000061