Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 1361–1366

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Studies on Antimalarial Substituted Phenyl Analogues and Their N^{ω}-Oxides

Vijay K. Agrawal,[a] Ruchi Sharma[a] and Padmakar V. Khadikar[b,*]

[a]*QSAR Laboratories, Department of Chemistry, A.P.S. University, Rewa 486 003, India*
[b]*Research Division, Laxmi Pest and Fumigation Pvt. Ltd. 3, Khatipura, Indore 452 007, India*

Received 25 June 2001; accepted 9 November 2001

Abstract—A quantitative structure–activity relationship (QSAR) study on a series of substituted phenyl analogues 5-[(7-chloro-4-quinolinyl)amino]-3-[(alkylamino)methyl](1,1′-biphenyl)-2-ols and their N^{ω}-oxides was made using various combinations of electronic and topological parameters. Several statistically significant regression expressions were obtained using multiple regression analyses. These regressions may be considered as mathematical models for investigating antimalarial activities of the compounds under present study. The antimalarial activity mechanism was investigated using combinations of E_L and E_H, independently with other molecular descriptors.

Introduction

Malaria, was and still is, recognized as a disease of world wide incidence, it is a highly wide-spread infectious disease caused by *soporoza of genus plasmodium*. It is characterized clinically by periodic fever, anemia, and enlargement of liver and spleen. It is the cause of higher sickness and death rate than any other disease, particularly in tropical and subtropical countries like Asia, Africa, and South America. Due to cooperation of all the countries and World Health Organization; this disease is controlled to a large extent. However, there appears to be a return of this disease because the malarial organisms have developed resistance against the drugs which are being used to control them. Consequently, there is consistant reporting of antimalarial drugs in the literature. The present paper deals with such a new series of antimalarials[1,2] and is an extension of our earlier work on topological modeling of antimalarials.[3–6] The results obtained will be helpful to pharmacologists, chemists, and medicinal chemists to synthesize still better anti-malarial drugs.

Quantitative structure–activity relationship (QSAR) method is the model proposed first by Hansch.[7] It was the seminal contribution to this field. The success of this method has prompted many workers to re-examine the derivation of Hansch equation by using the principles of theoretical pharmacology or pharmacokinetics.

QSAR method is also used to study the problem of pattern recognition. From this point of view, techniques of pattern recognition have been applied to QSAR studies to examine structural features and/or chemical properties underlying patterns that were associated with different biological effects.

Since the earliest days of QSAR calculations for drugs, frontier orbital energies, namely, the energy of highest occupied molecular orbital (E_H), the energy of lowest unoccupied molecular orbital (E_L), and total π-electron energy (E_T) have been used as electronic descriptors of drug molecules.[8,] The justification given for this has often been that E_H is a measure of the ability of the drug molecule to serve as an electron-donor in the formation of charge transfer complex (CTC) and, similarly E_L is a measure of the ability of the molecule to serve as an acceptor.[3]

Our earlier study on antimalarials[3–6] indicated that HMO method alone is not always useful in deriving final conclusions regarding modeling of antimalarial activity of organic compounds acting as drugs. The obvious reason being HMO is a simple semi-empirical quantum chemical method, the conclusions derived from such calculations may not be always correct. Hence, models based on E_H (energy of HOMO), E_L (energy of LUMO), E_T (energy of π-electrons), and CD (charge

*Corresponding author. Tel.: +91-731-531906; e-mail: vijay-agrawal@lycos.com

0968-0896/02/$ - see front matter
PII: S0968-0896(01)00399-6

density) alone may not be statistically excellent models. When these quantum chemical descriptors are combined with topological indices, excellent models are obtained. Furthermore, our earlier study[3–6] indicated that combination of Wiener[9] (*W*), Balaban[10] (*J*), and Szeged[11,12] (*Sz*) indices with some of the quantum chemical descriptors resulted in the statistically most significant models. Hence, our study deals with the combination of E_H, E_L, and *CD* on nitrogen atom in the side chain, with *W* and *Sz* in developing QSAR models for a series of antimalarials presented in Table 1 and Figure 1. The results, as discussed below, have shown that such combinations result in statistically significant models which are quite successful in estimating, modeling, and monitoring antimalarial activities of the compounds used in the present study.

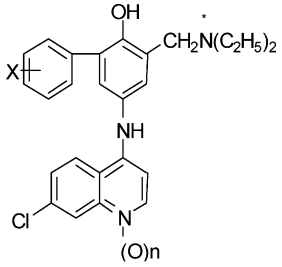

Figure 1. The antimalarial derivatives used in the present study.

The ability of malaria parasite to acquire in some, still unknown, manner a resistance to most known drugs remains a major problem for chemotherapeutists. As a solution to this problem, Werbel et al.[1] have synthesized some new antimalarials represented in Figure 1. These antimalarials were tested against a normal drug-sensitive strain of *Plasmodium berghei* in mice. The negative logarithm of molar ED_{30} (Table 1), the dose that provided a 30-day extension in life span, was used as the measure of potency, was calculated and values ranged[2] from −1.872 to 1.68 with a standard error of 0.28. We have adopted these $-\log ED_{30}$, that is pED_{30} in the present investigation to establish QSAR for the aforementioned antimalarials.

Results and Discussion

A perusal of Table 2 shows that high degeneracy is present in molecular orbital parameters E_H, E_L, and *CD* but comparatively lower degeneracy is observed in *W*, *Sz*, *B* and *J*. In comparison to these molecular descriptors the degeneracy in the activity (Table 1) is very little. The observed degeneracy in the molecular descriptors is due to the fact that they belong to first-generation molecular descriptors.[15] Balaban has shown that in spite of the observed high degeneracy the first generation molecular descriptors give better QSAR/QSPR models.

Table 1. Structural details, activity (pED_{30}) and comparison of observed and estimated values of pED_{30} of the antimalarials used in the present study

Compd	X	*n*	Obs. pED_{30}	pED_{30} estimated from			
				Model 2		Model 4	
				Est.	Res.	Est.	Res.
1	2-Cl	0	1.480	1.371	0.109	1.334	0.146
2	2-Cl	1	1.240	1.479	−0.239	1.416	−0.176
3	2-OCH_3	0	0.800	0.737	0.063	0.629	0.171
4	2-OCH_3	1	0.450	0.845	−0.395	0.715	−0.265
5	2,3-$(CHCH)_2$	0	0.940	Outlier	—	Outlier	—
6	2,5$(OCH_3)_2$	1	−0.680	−0.662	−0.018	−0.742	0.062
7	3-CF_3	0	1.480	0.833	0.647	0.782	0.696
8	3-CF_3	1	0.660	0.940	−0.280	0.874	−0.214
9	3,4$(OCH_3)_2$	0	−0.280	−0.838	0.558	−0.757	0.477
10	3,4,Cl_2	0	1.340	1.474	−0.134	1.492	−0.152
11	3,4,Cl_2	1	1.150	1.581	−0.431	1.577	−0.427
12	4-Cl	0	1.870	1.367	0.503	1.418	0.452
13	4-Cl	1	1.720	1.474	0.246	1.498	0.222
14	4-F	1	1.370	1.474	−0.104	1.498	−0.128
15	4-(OCH_3)	0	0.610	0.732	−0.122	0.793	−0.183
16	2-CF_3	0	1.160	0.909	0.251	1.127	0.033
17	2-CF_3	1	0.960	1.047	−0.087	1.124	−0.164
18	2-F	0	1.220	1.367	−0.147	1.418	−0.199
19	2-F	1	0.980	1.479	−0.499	1.416	−0.436
20	2,3,4,5,6-F_5	0	1.850	1.806	0.044	1.598	0.252
21	3-F	0	1.560	1.367	0.193	1.375	0.185
22	3-F	1	1.140	1.474	−0.334	1.455	−0.315
23	4-F	0	1.710	1.367	0.343	1.418	0.292
24	H	0	1.620	1.263	0.357	1.295	0.325
25	H	1	1.230	1.370	−0.140	1.371	−0.141
26	4-CF_3	0	1.710	1.684	0.026	1.847	−0.137
27	4-CF_3	1	1.840	1.732	0.108	1.883	−0.043
28	4-CH_3	1	−1.680	−1.097	−0.583	−1.100	−0.580
29	2,3-F_2	0	1.730	1.479	0.251	1.411	0.319
30	2,6-F_2	0	1.290	1.479	−0.189	1.365	−0.075

Res., difference between observed and estimated pED_{30}.

The same is found to be true in the present case also. The details are given below.

At this stage, it is interesting to record that we can use the aforementioned molecular descriptors to explain the possible mechanism of action of antimalarials under the present study. Recently, the antimalarial action mechanism of nanaomycin A and radicicol has been revealed, showing that the metabolic mechanism of these two compounds is similar to the peroxide compounds, acting with heime through radical reaction. If we believe that the compounds of phenol analogous under present study go along with the similar mechanism of nanaomycin A and radicicol then their activity should correlate with the energy of LUMO (i.e., E_L), for the electron of heime transfer to the drugs, forming a series of radicals to kill the plasmodium. Hence, in multiple linear analysis (MLA), other parameters remaining the same, multiparametric model containing E_L as one of the correlating parameters, should be statistically more significant than the models in which E_H is used in place of E_L. Whether this is so or not we have carried out simple and multiple linear regression analysis. The models obtained are discussed below.

Table 2. Topological indices calculated for compounds presented in Table 1

Compd	*W*	*B*	*J*	*Sz*	E_H	E_L	*CD*
1	3724	16.4603	1.5961	5168	−0.24	−0.3983	0.320
2	3976	16.8709	1.6214	5482	−0.24	−0.3983	0.320
3	3994	16.9983	1.6212	5534	−0.5	−0.3983	0.397
4	4258	17.4090	1.6454	5863	−0.25	−0.3983	0.397
5	4867	18.3408	1.6889	6738	−0.21	−0.2452	0.470
6	4648	17.6548	1.6396	6488	−0.24	−0.3883	0.398
7	4940	18.0654	1.6625	6855	−0.24	−0.3883	0.398
8	4678	17.9301	1.6305	6576	−0.19	−0.2441	0.476
9	4044	16.8541	1.6012	5666	−0.24	−0.3982	0.320
10	4310	17.2648	1.6258	6000	−0.24	−0.3982	0.320
11	3778	16.4434	1.5741	5276	−0.24	−0.3983	0.320
12	4032	16.8541	1.5999	5594	−0.24	−0.3983	0.320
13	4032	16.8541	1.5999	5594	−0.24	−0.3982	0.320
14	4102	16.9814	1.5803	5750	−0.25	−0.3983	0.397
15	4756	17.6548	1.5741	5276	−0.25	−0.3983	0.397
16	5052	18.0654	1.6274	7079	−0.25	−0.3983	0.394
17	3778	16.4434	1.5741	5276	−0.24	−0.3983	0.320
19	3976	16.8709	1.6214	5482	−0.24	−0.3983	0.320
20	4814	18.1198	1.7140	6768	−0.24	−0.3983	0.320
21	3751	16.4434	1.5846	5222	−0.24	−0.3983	0.320
22	4004	16.8541	1.6101	5538	−0.24	−0.3983	0.320
22	3778	16.4434	1.5741	5276	−0.24	−0.3983	0.320
23	3488	16.0496	1.5605	4836	−0.24	−0.3982	0.320
24	3729	16.4603	1.5871	5136	−0.24	−0.3982	0.320
25	4756	17.6548	1.6039	6704	−0.24	−0.3982	0.320
26	5052	18.0654	1.6274	7079	−0.24	−0.3892	0.320
27	4032	16.8541	1.5997	5594	−0.10	−0.2475	0.476
28	3990	16.8709	1.6226	5558	−0.24	−0.3983	0.320
29	3964	16.8709	1.6335	5506	−0.24	−0.3983	0.320

The first step in obtaining a statistically significant model is to investigate whether or not any collinearity exists between the parameters used. This is achieved by obtaining correlation matrix. Such a matrix obtained in the present case is shown in Table 3. In practice, every term in the correlation matrix >0.4 can be taken as being suspicious due to collinearity. Thus, this Table 3 contains important information that *W* and *B* are significantly correlated and that appreciable collinearity exists between *Sz*, *W*, *B*, and *J*. The data presented in Table 3 also show that pED_{30} can be modeled by using E_L and *CD* as the correlating parameters. The question of where to draw the line concerning collinearity is a controversial one. One should not be too harsh on oneself in deciding the problem of collinearity. We do not think that two descriptors that are only 'half' correlated can be considered as collinear.

Table 3. Correlation matrix for the correlation of pED_{30} of antimalarials presented in Table 1 using correlation parameters E_H, E_L, E_T, and *CD*

	pED_{30}	*W*	*B*	*J*	*Sz*	E_H	E_L	*CD*
pED_{30}	1.0000							
W	0.1867	1.0000						
B	0.2739	0.9756	1.0000					
J	0.2264	0.6424	0.7583	1.0000				
Sz	−0.0053	0.5183	0.5145	0.5268	1.0000			
E_H	0.1146	−0.0481	−0.0443	−0.1845	−0.8561	1.0000		
E_L	0.8441	0.2656	0.3483	0.2784	0.2076	−0.0086	1.0000	
CD	0.8643	0.4552	0.5215	0.3070	0.0703	0.1590	0.7945	1.0000

E_H, energy of highest occupied molecular orbital; E_L = energy of lowest unoccupied molecular orbital; these energies: E_H, and E_L are in β units; *CD* = Charge density on nitrogen atom in the side chain; pED_{30} is the activity.

Table 4. Proposed models for modeling pED_{30} of the compounds under present study (See Table 1 for details)

I. Models based on the combinations of E_L with other molecular descriptors

1. $pED_{30} = 7.1080(\pm 2.2370)E_L + 9.1432(\pm 2.0730)CD - 2.1843\times10^{-4}(\pm 1.0625\times10^{-4})Sz - 0.3445$
2. $pED_{30} = 6.6370(\pm 2.2264)E_L + 10.0686(\pm 2.1807)CD - 0.2622(\pm 0.1176)B + 2.3670$
3. $pED_{30} = 6.3122(\pm 2.2080)E_L + 10.1476(\pm 2.1315)CD - 3.7493\times10^{-4}(\pm 1.5356\times10^{-4})W - 0.6952$
4. $pED_{30} = 5.8747(\pm 2.2333)E_L + 10.9781(\pm 2.2150)CD + 4.1990(\pm 2.8172)J - 0.4573(\pm 0.1745)B - 1.6826$

II. Models based on the combinations of E_H with other molecular descriptors

5. $pED_{30} = -0.0015(\pm 8.4465\times10^{-4})E_H + 14.3406(\pm 1.5779)CD - 2.3454\times10^{-4}(\pm 1.2597\times10^{-4})Sz - 4.4996$
6. $pED_{30} = -2.1953\times10^{-4}(\pm 3.9269\times10^{-4})E_H + 15.2434(\pm 1.6578)CD - 0.3182(\pm 0.1364)B - 1.0324$
7. $pED_{30} = -2.2102\times10^{-4}(\pm 3.8138\times10^{-4})E_H + 15.1069(\pm 1.5426)CD - 4.6658\times10^{-4}(\pm 1.7450\times10^{-4})W - 4.4669$
8. $pED_{30} = -7.6273\times10^{-5}(\pm 3.8405\times10^{-4})E_H + 15.5804(\pm 1.5474)CD + 5.7664(\pm 3.1780)J - 0.5716(\pm 0.1912)B - 6.1232$

III. Models containing both E_H and E_L with other molecular descriptors

9. $pED_{30} = -2.2449\times10^{-5}(\pm 3.6118\times10^{-4})E_H + 10.4526(\pm 2.3980)CD + 6.3378(\pm 2.3820)E_L + 4.4470(\pm 2.5817)J - 0.4538(\pm 0.1784)B - 1.7873$

We have carried out regression analysis first by considering the combinations of E_L with other parameters and secondly using combination of E_H with other analogous parameters. The results are shown in Table 4.

In view of the above, and in an attempt of proposing possible action mechanism, we have carried out regression analysis[13,15,16] on the basis of combinations of E_L with other topological indices used in the present study and observed that tri-parametric correlations start giving statistically significant models (models 1–4 in Table 4). Out of the three tri-parametric models the model 3 gave the best results (Table 5). This model is found as:

$$pED_{30} = 6.3122(\pm 2.2080)E_L + 10.1476 \times (\pm 2.1315)CD - 3.7493 \times 10^{-4} \times \left(\pm 1.5356 \times 10^{-4}\right)W - 0.6952 \quad (1)$$

The data presented in Tables 4 and 5 show that out of the three distance-based topological indices (*W, Sz*, and *B*), *W* is the most appropriate index with which E_L can be combined to obtain a statistically significant model. *CD* is another useful parameter for this purpose. The coefficients of each of these terms, namely E_L, *CD* and *W* involved in the above model indicate that *CD* and E_L play a positive role in modeling pED_{30} and that *W* has a retarding effect on the exhibition of activity, that is pED_{30}.

During the process of step-wise regression, we obtained several tetra-parametric regressions, out of which the model based on the combination of E_L, *CD*, *J*, and *B* gave the best results. This model is found as:

$$pED_{30} = 5.8747(\pm 2.2150)E_L + 10.9781 \times (\pm 2.2150)CD + 4.1990(\pm 2.8172)J - 0.4573(\pm 0.1741)\mathrm{B} - 1.6826 \quad (2)$$

This shows that in higher-parametric regressions *J* and *B* start playing dominant roles as compared with other distance-based topological indices used. Also, that E_L, *CD*, and *J* play positive and dominant roles in exhibiting pED_{30}.

The data presented in Table 5 show that this tetra-parametric model (model 4 in Table 5) gives slightly better results in comparison to the tri-parametric model (model 3 in Table 5). The values of quality factor[14] *Q* also favors this finding.

It is worth mentioning that no other higher parametric regression resulted in better models than those discussed above.

Since we are interested in using combinations of E_L, and E_H independently in explaining the possible action mechanisms of antimalarials under present study, we tried the regression analysis in that now E_H was combined with other topological indices in accordance with the combinations of E_L in developing models 1–4 (Table 5). Such analogous models in terms of E_H are given in Table 5 (models 7 and 8).

A perusal of Table 5 shows that out of the three tri-parametric models (5–7), model 7 is found to be the best. We record that this model is similar to model 3 (Table 5) in that E_H is replaced by E_L.

In case of E_H, the tetra-parametric model containing E_H, *CD, J*, and *B* resulted in the most appropriate model.

We have also tried penta-parametric models in which both E_H and E_L are present simultaneously. One such model based on the combination of E_H, E_L, *CD, B,* and *J* is shown in Table 4 (model 9). However, the data presented in Table 5 show that this model is statistically less significant than the tetra-parametric models discussed above (models 2 and 4, Table 5).

The aforementioned results and discussion show that out of the nine proposed models, models 1–4 are solely based on the combination of E_L with other topological indices. Analogously, the models 5–8 are solely based on the combination of E_H with similar combinations of topological indices. Hence, other parameters remaining the same, models 1–4 and 5–8 accounts for the combinations of E_L and E_H, respectively.

In order to confirm our results, we have estimated the values of pED_{30} using models 2 and 4 and correlated each of them with the observed values of pED_{30}. Such

Table 5. Regression parameters and quality of model proposed in Table 4

Model	*Se*	R_A^2	*R*	*F*	*P*	*Q*
1	0.3383	0.8220	0.9171	44.109	3.919×10^{-10}	2.7110
2	0.3340	0.8265	0.9193	45.455	2.862×10^{-10}	2.7520
3	0.3286	0.8320	0.9220	47.223	1.955×10^{-10}	2.8060
4	0.3261	0.8346	0.9264	36.312	7.469×10^{-10}	2.8410
5	0.3996	0.7516	0.5822	29.238	2.445×10^{-8}	2.2280
6	0.3864	0.7677	0.5903	31.845	1.066×10^{-8}	2.3040
7	0.3760	0.7800	0.5964	34.098	5.422×10^{-9}	2.3840
8	0.3699	0.7872	0.9042	26.897	1.465×10^{-8}	2.4440
9	0.3345	0.8259	0.9224	34.218	1.363×10^{-9}	2.7580

correlations as demonstrated in Figures 2 and 3, confirm our findings.

The regression parameters and quality of correlations shown in Table 5 indicate that models based on the combinations of E_L are of better quality than the analogous combinations of E_H. Hence, in accordance with our earlier statement, the mechanism of action of the antimalarials under the present study is similar to peroxide compounds.

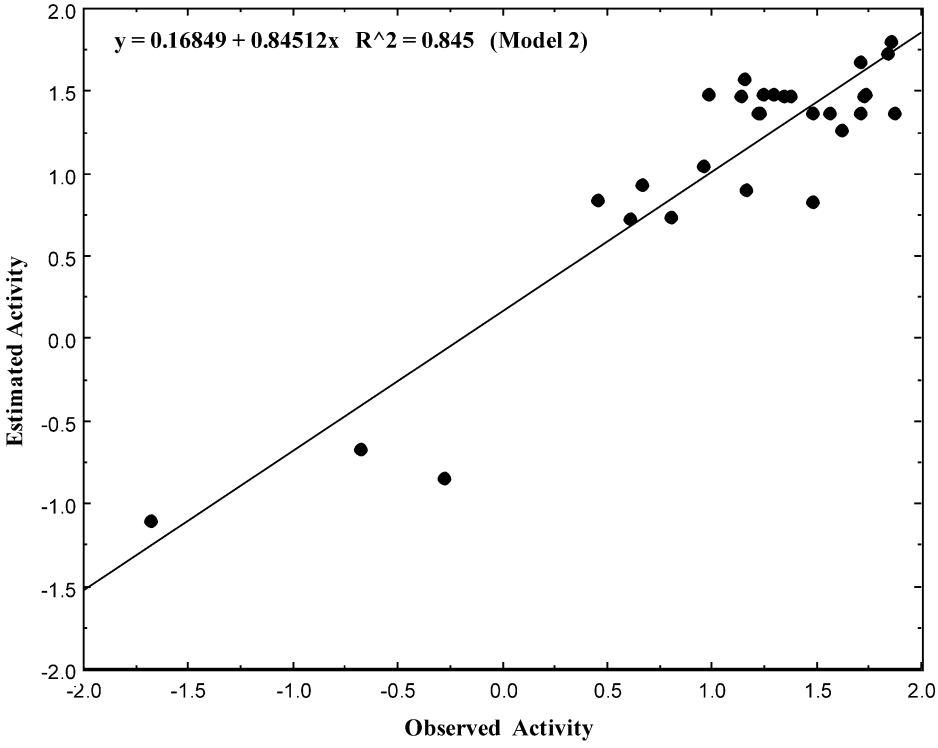

Figure 2. Correlation of observed pED_{30} estimated using model 2.

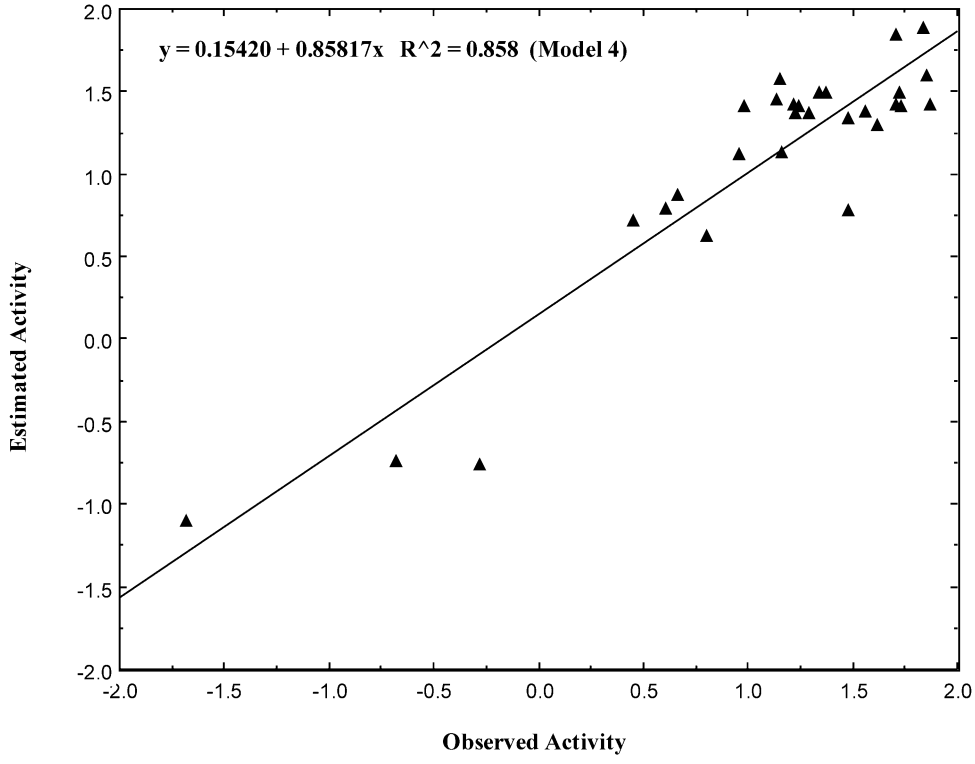

Figure 3. Correlation of observed pED_{30} estimated using model 4.

Conclusions

From the aforementioned results and discussion we conclude that the combinations of E_L with other molecular descriptors resulted in better quality models compared to those in which analogous E_H is involved. The results further suggest that the antimalarial action mechanism of the compounds used is similar to the peroxide compounds.

Experimental

Biological activity (pED_{30})

The biological activity, that is antimalarial activity as calculated and reported by Werbel[1] in terms of ED_{30} was converted into pED_{30}, ($-\log ED_{30}$) and used in the present investigation.

Calculation of electronic parameters

The electronic parameters, namely energy of the highest occupied molecular orbitals (E_H), energy of the lowest unoccupied molecular orbitals (E_L), the total π-electron energy (E_T) and charge density (CD) were estimated by HMO method.[3–6] All these energies are in β units.[2]

Calculations of topological indices

All the distance-based topological indices namely: W, Sz, B, and J were calculated using the program supplied by Prof. Istvan Lukovits.

Statistics

We have used the maximum R^2 improvement and cross-validation methods to identify predicted models. All regression models were examined by a variety of statistics associated with residues, that is the Wilks–Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results.[13]

Acknowledgements

Authors are thankful to Prof. A. Wissner and Prof. I. Lukovits for providing computer programs and to honorable referees for their excellent suggestions improving the quality of the paper.

References and Notes

1. Werbel, L. M.; Hung, J.; McNamara, D.; Ortwinc, D. F. *Eur. J. Med. Chem.* **1985**, *20*, 363. Werbel, L. M.; Dan Cook, P.; Elslager, E. F.; Hung, J. H.; Johnson, J. L.; Kesten, S. J.; McNamara, D. J.; Ortwine, D. F.; Worth, D. F *Med. Chem.* **1986**, *29*, 924.
2. Detailed computer output is available with one of the authors (P.V.K.) on request.
3. Sinha, S.; Bano, S.; Agrawal, V. K.; Khadikar, P. V. *Oxid. Commun.* **1999**, *22*, 479.
4. Agrawal, V. K.; Sinha, S.; Bano, S.; Khadikar, P. V. *Pol. J. Pharmacol.* **1999**, *51*, 517.
5. Agrawal, V. K.; Sinha, S.; Bano, S.; Khadikar, P. V. *Acta Microbiol.* **2001**, *48*, 17.
6. Agrawal, V. K.; Sinha, S.; Khadikar, P. V. *Indian J. Chem.* **2000**, *39B*, 472.
7. Hansch, C.; Leo, A. *Explaining QSAR. Fundamentals and Applications in Chemistry and Biology*; American Chemical Society, Washington, DC, 1995; p 65.
8. Franke, R. *Theoretical Drug Design Methods*; Elsevier: New York, 1984.
9. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
10. Balaban, A. T., Motoc, I.; Bonchev, D.; Mekenyan, O. In: *Steric Effects in Drug Design*; Charton, M., Motoc, I., Eds.; Akademic: Berlin, 1983; p 11.
11. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
12. Khadikar, P. V.; Deshpandey, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
13. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples,* 3rd ed; Wiley: New York, 2000.
14. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
15. Norrington, F. E.; Hyde, R. M. *J. Med. Chem.* **1975**, *18*, 604.
16. Groebner, D. F.; Shannon, P. W. *Business Statistics, a Decision Making Approach,* 3rd ed; Merrill: New York, 1999.