Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 3287–3293

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Studies on Some Antimalarial Sulfonamides

Vijay K. Agrawal,[a] Ravindra Srivastava[a] and Padmakar V. Khadikar[b,*]

[a]*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 16 April 2001; accepted 30 June 2001

Abstract—Antimalarial activity of a series of sulfonamide derivatives (2,4-diamino-6-quinazoline sulfonamides) was modeled topologically using Wiener (*W*)-, and Szeged (*Sz*)-indices. The regression analysis of the data has shown that better results are obtained in multiparametric regressions upon introduction of indicator parameters. Predicting ability of the models was tested by r^2_{cv} values. It was observed that models based on *W* index gave slightly better results than those in which *Sz* is involved.

Introduction

Quantitative structure–activity relationships (QSAR) are a major factor in contemporary drug designing. Thus, it is quite clear why a large number of users of QSAR are located in industrial research units. The coupling of QSAR methodology to the experience in intuition of professional drug designer should result in a much more organized search for the novel drugs of human, animal, and plant therapy.[1,2]

There are two fundamental kinds of molecular prediction used in QSAR studies. One of them involves parameters that bare relation to free energy and the other uses topological index or numerical graph invariants, which are produce directly from molecular structure.[3,4]

The interest in the influence of molecular topology on molecular activity has grown remarkably during the last few years. Though many numerical invariants have been devised so far, only a handful of them have been employed in QSAR studies.[5–7] A widely used index in these studies is the Wiener index[8] symbolized as *W*.

Gutman[9] and one of the present authors (PVK)[10] have introduced a novel topological index which eventually was named the Szeged index and denoted by *Sz*. This index, *Sz*, is a modification of Wiener index to cyclic graphs.

A number of successful QSAR studies were made based on the Wiener index[11] however, very little is known on the use of Szeged index in developing (QSAR).

Sulfonamides and its derivatives have been widely used in medicinal chemistry as antimalarials.[12] Of late, Elslager et al.[13] have tested 2,4-diamino-6-quinazoline sulfonamides for their antimalarial potentials. However, to date no topological QSAR modelling on this class of compounds has been reported in the literature.

In view of the above, we have chosen both *W*, and *Sz* indices by means of which we tried to propose QSAR models for modeling antimalarial properties of 2,4-diamino-6-quinazoline sulfonamides (Fig. 1, Table 1). In this respect, both *W*, and *Sz* were found quite successful. We further observed that better results are obtained in multiparametric regressions upon introducing indicator parameters in the regression analysis. The details are given below.

Results and Discussion

A perusal of Table 1 shows that degeneracy is present both in topological indices (*W*, *Sz*) as well as in antimalarial activity ΔMST. The degeneracy in *W* and *Sz* is obvious as they belong to first generation topological indices defined by Balaban.[14] According to Balaban[14] the first generation indices in-spite of their observed

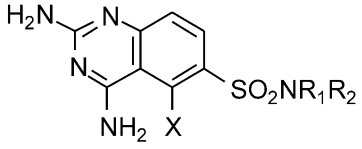

Figure 1. 2,4-Diamino-6-quinazoline sulfonamides used in the present study.

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

0968-0896/01/$ - see front matter
PII: S0968-0896(01)00241-3

Table 1. 2,4-diamino-6-quinazoline sulfonamides, their antimalarial activity, Wiener (*W*)-, and Szeged indices(*Sz*)-, and indicator parameters (Ip_1, Ip_2 and Ip_3)

	Structure		Property	Parameters					
S.N.	$-NR_1R_2$	*X*	ΔMST	*W*	*Sz*	$^1\chi^v$	Ip_1	Ip_2	Ip_3
1.	$N(C_2H_5)_2$	H	3.30	1240	2025	6.2265	1	0	0
2.	$N(CH_2)_5$	Cl	2.30	1370	2240	7.1679	1	0	0
3.	$N(CH_2\ CH_2\ CH_3)_2$	H	0.30	1575	2475	7.2265	1	0	0
4.	$N(CH_2\ CH_2\ OH)_2$	H	0.30	1982	2995	6.4447	0	0	1
5.	$N(CH_3)CH\ (CH_3)_2$	H	0.70	1218	2023	6.0399	1	0	0
6.	$N(CH_3)CH_2CH_2N(C_2H_5)_2$	H	0.10	2065	3078	7.8062	0	0	1
7.	$N(CH_2)_5$	H	4.40	1448	2458	6.8123	0	0	0
8.	$N(CH_2)_4$	H	5.00	1336	2141	6.3123	0	0	0
9.	$N[(CH_2)_2]_2O$	H	4.70	1448	2458	6.3896	0	0	1
10.	$N[(CH_2)_2]_2S$	H	2.50	1448	2458	6.3896	0	0	1
11.	$N[(CH_2)_2]_2NCH_3$	H	1.00	1585	2750	6.8920	0	0	1
12.	$N[(CH_2)]_2NC(=O)OC_2H_5$	H	0.20	2518	4173	7.7795	0	0	1
13.	$4Cl\text{-}C_6H_4NH$	H	0.70	1971	3272	6.4617	0	1	0
14.	$3Br\text{-}C_6H_4NH$	H	0.30	1950	3230	6.4617	0	1	0
15.	$4Cl\text{-}C_6H_4NCH_3$	H	0.30	1950	3230	6.3316	0	1	0
16.	$C_6H_5NCH_3$	H	0.30	1756	2928	6.7613	0	1	0

degeneracy are quite successful in developing statistically significant QSAR models. The present study also shows that *W* and *Sz* indices can be used successfully in modeling antimalarial activity of the compounds used.

An inspection of Table 1 shows that *W* and *Sz* are moderately related to antimalarial activity. In both the cases negative correlation were observed indicating that the activity increases with decrease in magnitude of both *W* and *Sz*.

The data presented in Table 1 show that the antimalarial activity (ΔMST) follow the following sequence:

$$8 > 9 > 7 > 1 > 10 > 2 > 11 > 5 = 13 > 16 > 3 = 4 = 14 = 15 > 12 > 6$$

The aforementioned sequence, however, does not exhibit any QSAR. The sequence also indicates that no statistical significant mono-parametric correlations are possible for modeling the activity. It means that antimalarial activity could better be modelled by multivariate analysis.

In view of the above we have carried out several multivariate correlations and observed that the analysis gave moderate to significant QSAR models (Table 3).

The first step in analyzing multivariate correlation is to investigate auto-correlation. This can be achieved by obtaining correlation matrix.[16,17] The correlation matrix is useful in determining which independent variables are likely to help explain variation of dependent variable (antimalarial activity in present case). We look for correlations close to ± 1.0. Such a correlation matrix obtained in the present case is given in Table 2.

A perusal of Table 2 indicates that *W* and *Sz* indices are highly correlated. Therefore, the regression involving these topological indices together will suffer from the defect of auto-collinearity and such models will give redundant information.

Inspection of Table 3 also shows that *W* and *Sz* indices each are moderately correlated with antimalarial activity. None of the other molecular descriptors show any auto-correlation and none of them correlate with the activity.

The above results, therefore, indicate that multiple correlation involving *W* or *Sz* as one of the correlating parameters will give statistically significant models for modeling the activity.

Table 2. (a) Correlation matrix for inter-correlation of molecular descriptors and their correlation with antimalarial activity (ΔMST). (b) Changed correlation matrix due to additional use of $^1\chi^v$ for inter-correlation of molecular descriptors and their correlation with antimalarial activity (ΔMST)

(a)	*W*	*Sz*	ΔMST	Ip_1	Ip_2	Ip_3
W	1.0000					
Sz	0.9787	1.0000				
ΔMST	−0.6501	−0.6094	1.0000			
Ip_1	−0.5385	−0.5748	−0.0042	1.0000		
Ip_2	0.3739	0.4340	−0.4072	−0.3333	1.0000	
Ip_3	0.3577	0.3327	−0.0882	−0.4472	−0.4472	1.0000

(b)	*W*	*Sz*	$^1\chi^v$	ΔMST	Ip_1	Ip_2	Ip_3
W	1.0000						
Sz	0.9787	1.0000					
$^1\chi^v$	0.5416	0.4910	1.0000				
ΔMST	−0.6501	−0.6094	−0.3421	1.00000			
Ip_1	−0.5382	−0.5795	−0.0602	−0.0042	1.0000		
Ip_2	0.3741	0.4340	−0.2407	−0.4072	−0.3333	1.0000	
Ip_3	0.3572	0.3327	0.3476	−0.0882	−0.4472	−0.4472	1.0000

At this stage it is worth recording that multiple regression analysis generally requires sufficiently more compounds than parameters; a useful rule of thumb is three to six times the number of parameters under consideration. Hence, traditional regression methods require that the number of parameters must be considerably smaller than the number of compounds in the data set (or the number of degrees of freedom data). The lower limit being three times as many molecules as independent variables[15] (parameters/descriptors). As will be seen below, in the present study the best model is tetra-parametric consisting of four descriptors for the set of sixteen compounds. This is in accordance with slightly more than the accepted lower limit of the referred thumbs rule.

The results of simple and multiple correlations are given in Table 4. Out of several correlations attempted by us only ten were found to give statistically moderate to significant results.

Table 4 shows that simple correlations involving W and Sz gave moderate correlations with the activity. Therefore, for mono-parametric regression these are not good topological indices. Consequently, we have attempted several multi-parametric regressions for obtaining better representation of the molecules. We observed that only eight statistical good multiple correlations were obtainable.

In the correlations presented in Table 3 the coefficients of W and Sz terms are negative. This indicates that the antimalarial activity increases with the decrease magnitude of both W and Sz. Because, W and Sz represent size and bulk of the molecule, the negative signs of these coefficients indicate that shape and size of the antimalarials have a opposite effect on the activity.

A close look of Table 4 shows that in mono-, bi-, tri-, and tetra-parametric correlations, the correlations involving W gave slightly better results than those in which Sz is involved. The tetra-parametric correlation involving W, $^{I}p_1$, $^{I}p_2$ and $^{I}p_3$ is found most appropriate for modelling antimalarial activity in accordance with the following equation:

$$\Delta MST = -0.0032\ (\pm 0.0010)W - 3.1824\ (\pm 0.9143)^{I}p_1 - 2.5978\ (\pm 1.0591)^{I}p_2 - 1.7911(\pm 0.9807)^{I}p_3 + 9.1679 \quad (2)$$

It is worthy of mention that the sign of the coefficient of all the three indicator parameters in all the multiple correlations is negative. Recall that we have used $^{I}p_1$ to account for the substitution of one of the hydrogen atoms of the amino group by an acyl group, therefore, increasing acyl chain will have a retarding effect on the antimalarial activity. Similarly, when aromatic ring is present the indicator parameter is represented by $^{I}p_2$. The above results, therefore, indicate that the presence of aromatic ring in the side-chain substituent decreases the antimalarial activity. The retarding effects due to these indicator parameters is found to be similar.

The indicator parameter $^{I}p_3$ is used for representing the presence of two or more hetero-atoms in the molecule. The negative sign of this parameter in the aforementioned correlation also indicates retarding effect due to the presence of multi-hetero-atom system. Surprisingly, the retarding effect of $^{I}p_3$ is also found similar to that of $^{I}p_1$ and $^{I}p_2$.

The quality of regression models presented in Table 3 are given in Table 4. Therefore, using this Table and

Table 3. Statistically significant QSAR models for modelling antimalarial activity

Model No.	Regression expression
(1)	$\Delta MST = -0.0032\ (\pm 9.9221\times10^{-4})W + 6.9977$
(2)	$\Delta MST = -0.0019\ (\pm 6.3548\times10^{-4})Sz + 6.8222$
(3)	$\Delta MST = -0.0045\ (\pm 0.0010)W - 1.9814(\pm 0.08269)^{I}p_1 + 9.6975$
(4)	$\Delta MST = -0.0028\ (\pm 6.9413\times10^{-4})Sz - 2.1021(\pm 0.8938)^{I}p_1 + 9.9246$
(5)	$\Delta MST = -0.0041\ (\pm 0.0010)W - 2.1844\ (\pm 0.8013)^{I}p_1 - 1.0922(\pm 0.7280)^{I}p_2 + 9.4033$
(6)	$\Delta MST = -0.0026\ (\pm 7.2234\times10^{-4})Sz - 2.2182\ (\pm 0.8888)^{I}p_1 - 0.9272(\pm 0.8072)^{I}p_2 + 9.4696$
(7)	$\Delta MST = -0.0019\ (\pm 7.5770\times10^{-4})Sz - 3.2559\ (\pm 0.9977)^{I}p_1 - 2.6109(\pm 1.911)^{I}p_2 - 1.9345\ (\pm 1.0718)^{I}p_3 + 9.0548$
(8)	$\Delta MST = -0.0032\ (\pm 0.0010\ )W - 3.1824\ (\pm 0.9143)^{I}p_1 - 2.5978\ (\pm 1.0591)^{I}p_2 - 1.7911\ (\pm 0.9807)^{I}p_3 + 9.1679$

Table 4. Quality of QSAR models recorded in Table 3

	Regression parameters					
Model no.	Se	R^2_A	R^2	R	Q	F
1	1.3963	—	0.4229	−0.6503	0.4657	10.258
2	1.4574	—	0.3713	−0.6093	0.4181	8.268
3	1.2068	0.5381	0.5997	0.7744	0.6417	9.737
4	1.2667	0.4911	0.5590	0.7476	0.5902	8.238
5	1.1527	0.5756	0.6629	0.8146	0.7063	7.866
6	1.2514	0.5033	0.6027	0.7763	0.6203	6.067
7	1.1481 (0.2913)	0.5820	0.6934	0.8327 (0.9890)	0.7253 (3.3951)	6.221
8	1.0546 (0.2426)	0.6473	0.7414	0.8610 (0.9924)	0.8164 (4.090)	7.882

Values in the paranthesis are obtained considering compounds **5**, **9**, and **11** as outliers.

considering correlation coefficient R alone as the deciding factor for the quality then the regression-8 will be the most significant model for modeling antimalarial activity. In addition to R the smallest standard error of estimation also favors eq 8.

The Q-values[18] ($=R/\mathrm{Se}$) for all the ten correlations are presented in Table 4 which also indicate that the correlation-8 is the most appropriate as it has the higher Q-value. The next higher value of Q is found for regression-7.

Note that both the regressions-7 and -8 are tetravariate correlations. In the regression-7 the correlating parameters involved being Sz, $I\mathrm{p}_1$, $I\mathrm{p}_2$ and $I\mathrm{p}_3$, while in regression-8 the parameters involved are W, $I\mathrm{p}_1$, $I\mathrm{p}_2$ and $I\mathrm{p}_3$. Except for W and Sz other three parameters are common in both correlations. The observed result, therefore, indicate that use of W in place of Sz results into better models. This is born-out from the values of adjustable R^2 (R^2_{A}) values. Note that R^2_{A} takes into account of adjustment of R^2. Therefore, if a variable is added that does not contribute its fare share then R^2_{A} will decline. In the present case the replacement of W for Sz resulted into increase in the value of R^2_{A}. In fact highest R^2_{A} value is observed for the regression-8 involving W, $I\mathrm{p}_1$, $I\mathrm{p}_2$ and $I\mathrm{p}_3$ molecular descriptors. Hence, the highest value of Q and R^2_{A} are in favor of this regression model.

As stated earlier, the methodology used in the present study fulfills slightly more than the lower limit of thumb rule required for the use of number of parameters (independent variables/descriptors) as compared to number of compounds (molecules) under study. The parameters used being topological indices (W, Sz) and indictor parameters ($I\mathrm{p}_1$, $I\mathrm{p}_2$, $I\mathrm{p}_3$). One can, therefore, think to describe for antimalarial activity only topological and indicator parameters is lack of parameter in the proposed model. Thus, possibly other descriptors like electronic and hydrophobic parameters are needed. We could not calculate hydrophobic parameter, but we did calculate electronic parameter, viz. $^1\chi^{\mathrm{v}}$, the widely used descriptor in QSPR/QSAR studies. The changed correlation matrix due to the addition of this descriptor is given at the lower part of Table 2. The changed correlation matrix shows that the contribution of $^1\chi^{\mathrm{v}}$ to the activity (ΔMST) is statistically insignificant. We tried several combinations of $^1\chi^{\mathrm{v}}$ with the indictor parameters used but none resulted into any statistically significant model. This indicates that the antimalarial activity of sulfonamides used is not controlled by this parameter. Further combination of $^1\chi^{\mathrm{v}}$ with W, Sz or both W and Sz along with indicator parameters resulted into models in that the standard error of $^1\chi^{\mathrm{v}}$ term was much higher than its coefficients. Such models are not allowed statistically.

As opposed to traditional regression models (multiple regression method used in the present study), cross-validation method evaluates the validity of a model by how well it predicts data rather than how well it fits data. The analysis uses a 'leave-one-out' scheme; a model is built with N-1 compounds and the Nth compound is predicted. Each compound is left out of the model derivation and predicted in tern. An indication of the performance of the model is obtained from cross-validation parameters. We have, therefore, calculated this cross-validation parameters as mentioned below.

Now obtaining most appropriate model does not mean that it will also have the highest predictive ability. The problem before us is, therefore, to decide the aptness of regression models-7 and -8. This can be achieved by obtaining cross-validation correlation coefficient r^2_{cv}. The closer this correlation[15] is to unity, the better the predictive ability. The values of this coefficient is determined by the following expression:

$$r^2_{\mathrm{cv}} = 1 - \frac{\mathrm{Press}}{\mathrm{SSY}}$$

PRESS and SSY values for all the multiple correlations attempted are given in Table 6. This Table also records the value of r^2_{cv}, which is found to be the highest for the regression-8. This means that the predictive ability of the model based on regression-8 is higher.

Further confirmation in favor of our results is obtained by determining uncertainty in the prediction[15] (S_{PRESS}). The uncertainty in the prediction (S_{PRESS}) is determined by the following expression:

$$S_{\mathrm{press}} = \sqrt{\frac{\mathrm{Press}}{N-K-1}}$$

where N is the number of samples and K is the number of variables involved in the model.

The values of N, K, $N-K-1$, and S_{PRESS} are recorded in the Table 6. The lowest value of S_{PRESS} for regression-8 justifies its predictive ability over others. These results, therefore, show that the most appropriate regression model and the highest predictive ability are two different aspects of regression analysis. The most appropriate model may or may not exhibit its predictive ability.

In order to confirm our results we have estimated antimalarial activity from regressions-7 and -8 (Table 3), compared them with their observed values. The plot of observed versus estimated activity gave R values of 0.8602 and 0.8683, respectively, for eqs (7) and (8). This shows that use of W gives slightly better results.

Further confirmation of the relative predictive ability of regression 7 and 8 is made by obtaining quality control chart (Fig. 2), which again favors that the predictive ability of regression 8 is better than 7.

Recall that Sz is a modification of W for cyclic graphs. In the set of antimalarials used, the basic unit in all the

cases contains two cycles (aromatic rings). Therefore, effect due to cycle (presence of aromatic ring) will be similar in all the cases. Any change in *Sz* and consequent changes in antimalarial activity should be attributed to changes in tree-like structures in the substituents for which the coincidence of *W* and *Sz* is well known.[9,10]

The aforementioned results indicate that compared to *Sz*, *W* is a better topological index for modelling antimalarial activity of sulfonamides used in the present study. These results also indicate that models based on regressions 7 and 8 (Table 3) gave large deviation of calculated activity values for the compounds **5**, **9**, and **11** which constitute about 20% of the total compounds used for obtaining the regression models. Upon omitting these compounds from the data set the quality of models expressed by eqs (7) and (8) is significantly increased to the extent that now for the most significant tetra-parametric correlations 7 and 8, *R*-values are found as 0.9890 and 0.9924, respectively. Consequent to the removal of these compounds the standard error of estimation (Se) is also decreased significantly and giving Se as 0.2913 and 0.2426, respectively, for the models 7 and 8.

Consequent to above, the quality factor *Q* for the models 7 and 8 are found as 3.3951 and 4.0907, respectively. However, removal of the aforementioned compounds from the data set does not change our final findings i.e., even after deletion of such compounds the tetra-parametric correlation involving *W*, $^{I}p_1$, $^{I}p_2$, and $^{I}p_3$ is found to be most appropriate than in which *Sz* is involved for modelling antimalarial activity. But this modified regression model should be used cautiously as now model is based on four variables with only thirteen data points[20,21] and may suffer from the defect of chance correlation. The results definitely indicate that antimalarial activity of the type of the compounds used, could be modelled successfully using the model based on regression 8. The model based on regression 7 would be slightly worsted for this purpose.

From the aforementioned results and discussion we arrived at the conclusion that the selection of the descriptors to be used in structure–property–activity (QSPR/QSAR) studies should not be delegated solely to the computers although the statistical criteria will continue to be useful for preliminary screening of descriptors taken from a large pool.

Conclusions

From the aforementioned results and discussion the following conclusions are made.

1. No univariate regression is possible for modelling antimalarial activity of the compounds used in the present study;
2. Step-wise regression method has indicated that addition of indicator parameters in the multivariate correlations gives better results.
3. In all the multiple regression models the coefficient of indicator parameters ($^{I}p_1$, $^{I}p_2$, $^{I}p_3$) is negative meaning thereby that they have retarding effect on the activity. This retarding effect is found to be similar for all the three indicator parameters;
4. The most statistically significant models were tetra-parametric expressions. Two such models involving: (1) *W*, $^{I}p_1$, $^{I}p_2$ and $^{I}p_3$, and (ii) *Sz*, $^{I}p_1$, $^{I}p_2$ and $^{I}p_3$ are found to be most relevant. Based on R^2_A and *R* values the former model was found to be most appropriate.
5. Cross-validation correlation coefficient (R^2_{cv}) as well as uncertainty in prediction (S_{PRESS}) have also indicated that regression involving *W*, $^{I}p_1$, $^{I}p_2$ and $^{I}p_3$ parameters is the most appropriate. The predicting ability of the tetravariate regression involving *Sz*, $^{I}p_1$, $^{I}p_2$ and $^{I}p_3$ is slightly worsted. This is further confirmed from the quality control charts.
6. Since the basic unit of the antiamlarials contains only two cycles, the variation in *Sz* is due only to tree-like structure present in the side-chain (-NR_1R_2) i.e., substituents, for which coincidence of *W* and *Sz* is well established. This is the reason why the correlations involving *W* is better than the correlation involving *Sz*.

Experimental

Material and methods

Pharmacology. In the present study we have adopted the antimalarial activity (ΔMST) reported by Elslager et al.[13] ΔMST is defined as the mean survival time (in days) of treated mice minus the mean survival time (in days) of controlled mice. For details see ref 13.

Wiener index(*W*). The Wiener index (*W*) is the oldest and widely used topological index.[7–10] It is based on the vertex-distances of the respective molecular graph.

Let us denote a molecular graph by *G* and having v_1, v_2, v_3,...,v_n its vertices. Let $d(v_i, v_j \mid G)$ stand for the distance between the vertices v_i and v_j. Then the Wiener index is defined as:

$$W = W(G) = 1/2\sum_{i=1}^{n}\sum_{j=1}^{n} d(v_i, v_j \mid G) \quad (1)$$

Szeged index (*Sz*). Let *e* be an edge of the molecular graph *G*. Let $n_1(e \mid G)$ be the number of vertices of *G* lying closer to one end of *e*; let $n_2(e \mid G)$ be the number of vertices of *G* lying closer to the other end of *e*. Then the Szeged index (*Sz*) is defined[9,10] as:

$$Sz(G) = Sz = \sum_{e} n_1(e \mid G) n_2(e \mid G) \quad (2)$$

with the summation giving over all edges of *G*.

In cyclic graphs, there are edges equidistant from both the ends of edge *e*; by definition of *Sz* such edges are not taken into account.

Indicator parameters (I_{p_1}, I_{p_2}, I_{p_3}). Indicator variables (parameters), sometimes called dummy variables or de novo constants, are used in multiple linear regression analysis to account for certain features which cannot be described by continuous variables. In QSAR equations they normally describe a certain structural element, be it a substituent or another molecular fragment. Thus, Free Wilson analysis may be interpreted as a regression analysis approach using only indicator variables.

The indicator parameters (variables) take on only two values, usually zero and one. The two values signify that the observation belongs in one of the two possible categories. The numerical values of the dummy variables are not intended to reflect a quantitative ordering of categories, but only serve to identify category or class membership. Therefore, they show the significance of a particular group or a substituent in a given series of drug. They account for the abrupt increase or decrease of a given pharmacological activity at any specific site in the drug molecule. If the coefficient of indicator parameter carries a negative sign in the regression expression, this makes it very clear that the compound having this particular group at a particular position will have considerable lower potency.

In the present case the indicator parameter I_{p_1} was taken as unity when one of the hydrogen of the NH_2 group is replaced by an acyl radical, in absence of such substitution I_{p_1} was zero. Similarly, I_{p_2} is used when benzene ring is present at the substituent, otherwise its value is taken as zero. When two or more hetero-atoms are present in the molecule the indicator parameter I_{p_3} is used having its value unity, otherwise it is zero.

Regression analysis. We have used the maximum R^2 improvement method to identify prediction models.[14,15] This method finds the 'best' one variable model, the 'best' two variable model and so forth for the prediction of property/activity. Several models (combinations of variables) were examined to identify combinations of variables with good prediction capabilities. In all regression models developed we have examined a variety of statistics associated with residues, i.e., the Wilks–Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results.[15,16] Finally, results are discussed on the basis of cross-validation parameters.

Multiple regression analyses for correlating antimalarial activities of the present set of compounds with the aforementioned molecular descriptors were carried out using Regress-1 software as supplied by Professor I. Lukovits, Hungarian Academy of Sciences, Budapest, Hungary. Several multiple regressions were attempted using correlation matrix from this program and the best results are considered and discussed in developing QSAR and hence, for modeling the antimalarial activities of the compounds in the present study.

Computations. All the computations were carried out in Power Macintosh 9600/233.

Presentation of data. 2,4-diamino-6-quinazoline sulfonamides used as antimalarials are presented in Table 1. This table also records calculated values of *W* and *Sz*, assumed indicator parameters (I_{p_1}, I_{p_2}, I_{p_3}) and antimalarial activity in terms of ΔMST.

The correlation matrix for the inter-correlation of the aforementioned parameters is given in Table 2.

The statistically significant correlations obtained for modeling antimalarial activities are presented in Table 3, while their quality of correlations are given in Table 4.

D-statistics for the significant models is given in Table 5 and the predictive-ability of the significant QSAR models is shown in Table 6.

Table 5. Durbin–Watson[19] treatment for the QSAR models 1 to 5 represented by regression eqs. 3 to 8 respectively

Model no.	Prameters	Samples	D-Statistics		
			D	dl	du
1	*W*, I_{p_1} (2)	16	1.4740	0.84	1.09
2	*W*, I_{p_1}, I_{p_2} (3)	16	1.3625	0.74	1.25
3	*Sz*, I_{p_1}, I_{p_2} (3)	16	1.3555	0.79	1.25
4	*Sz*, I_{p_1}, I_{p_2}, I_{p_3}, (4)	16	1.6298	0.63	1.44
5	*W*, I_{p_1}, I_{p_2}, I_{p_3}, (4)	16	1.4888	0.63	1.44

Table 6. Predicting ability of QSAR models 3,5,6,7 and 8

Model No.	Prameters	Samples	PRESS	SSY	PRESS/SSY	R^2_{cv}	N-K-1	S_{PRESS}
1	*W*, I_{p_1} (2)	16	18.9336	28.3639	0.6675	0.3325	13	1.2068
2	*W*, I_{p_1}, I_{p_2} (3)	16	15.9433	31.9433	0.4991	0.5009	12	1.1527
3	*Sz*, I_{p_1}, I_{p_2} (3)	16	18.9732	28.5043	0.6656	0.3343	12	1.2574
4	*Sz*,I_{p_1},I_{p_2},I_{p_3},(4)	16	14.4952	32.7983	0.4421	0.5579	11	1.1481
5	*W*,I_{p_1},I_{p_2},I_{p_3},(4)	16	12.2975	35.0641	0.3507	0.6493	11	1.0573

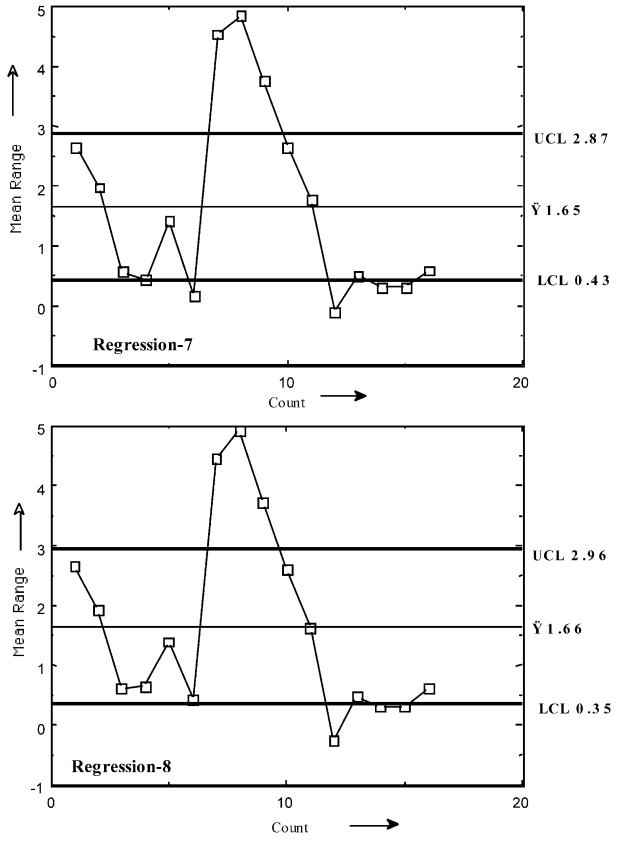

Figure 2. Quality control charts for forecasting antimalarial activity of the compounds in the present study using regressions 7, and 8, respectively.

The quality control charts for the most appropriate models are shown in Fig. 2.

Acknowledgements

The authors are thankful to Professor Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary, for providing programs calculating *W*, *Sz* and for carrying out regression analysis and to Professor Ivan Gutman for his consistent help and encouragement.

Authors also thank the honorable referees for their fruitful suggestions which improved the quality of this paper.

References and Notes

1. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry; Drug Research*; Academic Press: New York, 1976.
2. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Relationship*; Wiley: New York, 1986.
3. Trinajstic, N. *Chemical Graph Theory*, 2nd Ed.; CRC Press: Boca Raton FL, 1992.
4. Gutman, I.; Cyvin, S. J. *Introduction to the Theory of Benzenoid Hydrocarbons*; Spriger-Verlag: Berlin, 1989.
5. Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. *Indian J. Chem.* **1993**, *32A*, 651.
6. Trinajstic, N. *Chemical Graph Theory, Vol. I,II*; CRC Press: Boca Raton, FL, 1983.
7. King, R.B. Ed. *Chemical Applications of Topology; Graph Theory*; Elsevier: Amsterdam, 1983.
8. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
9. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
10. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
11. Saxena, A.; Khadikar, P. V. *Acta. Pharm* **1999**, *49*, 171.
12. Shao, B. R. *Chin. Med. J. Endl* **1990**, *103*, 428.
13. Elslager, E. F.; Colbry, N. L.; Davoll, J.; Hutt, M. P.; Johnson, J. L.; Werbel, L. M. *J. Med. Chem.* **1984**, *27*, 1740.
14. Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O.; in: *Steric Effects in Drug Design*; Charton, M.; Motoc, I. eds., Akademic-Verlag: Berlin, 1983.
15. Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
16. Shapiro, S. S.; Wilk, M. B. *Biometrica* **1988**, *52*, 591.
17. Cook, R. D. *Technometrics* **1977**, *19*, 15.
18. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
19. Durbin, J.; Watson, G. S. *Biometrica* **1951**, *38*, 159.
20. Topliss, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, *22*, 1238.
21. Topliss, J. G.; Costello, R. G. *J. Med. Chem.* **1972**, *15*, 1066.