

Application of topological and physicochemical descriptors: QSAR study of phenylamino-acridine derivatives

Abhilash Thakur^{a*}, Mamta Thakur^b, Nitika Kakani^c, Ashok Joshi^a, Suprajnya Thakur^d,
and Ashok Gupta^b

^a *Research Division, Upzon Drugs Pvt.Ltd. Sanwar Road Indore, India*

^b *Department of Chemistry, Softvision College of Biotechnology and Science, Indore, India*

^c *Department of Chemistry, New Girls PG College Kilamaidan, Indore, India*

^d *Department of Physics, Shree Vaishnave Institute of Technology & Science, Indore, India*

E-mail: abhilashthakur@yahoo.com

(received 31 May 04; accepted 09 Aug 04; published on the web 20 Aug 04)

Abstract

In the present investigation the applicability of various topological parameters are tested for the QSAR study on phenylamino-acridine derivatives. For the modeling of DNA binding affinity of phenylamino-acridine derivatives the regression analysis shows that even in the mono-parametric correlations the topological and physicochemical parameters give significant regression coefficients. Furthermore using combinations of topological, physicochemical parameters along with the indicator parameters, a tremendous improvement in the statistics has been observed. The results are critically discussed on the basis of regression data and cross-validation parameters.

Keywords: QSAR, topological descriptors, DNA binding, phenylamino-acridine

Introduction

Quantitative structure–activity relationship and Quantitative structure–property relationship (QSAR/QSPR) studies are indubitably of great importance in modern chemistry and biochemistry. To obtain a significant correlation, it is essential that appropriate descriptors are employed, whether they are theoretical, empirical or derived from readily available experimental characteristics of structures. Many descriptors reflect simple molecular properties and can thus provide insight into the physicochemical nature of the activity/ property under consideration.¹

To get an insight into the structure-activity relationship we need molecular descriptors that can effectively characterize molecular size, molecular branching or the variations in molecular shapes, and can influence the structure and its activities.²⁻⁷

If molecular structure is critical for understanding of a particular structure-activity and property-activity relationship, then one should consider structural invariants derived from molecular structure.⁸⁻¹⁹ Several graph theoretical invariants have been generalized so that they

produce structure-dependent descriptors.²⁰⁻²³ In such a generalization, most biological activities are dominated by molecular size, which is well characterized by most of the physicochemical properties.²⁴

Ideally, the activities and properties are connected by some known mathematical function, F: Biological activity = F [structure (in present study topological & physicochemical descriptors are used as the structural parameters.)]

Biological activity can be any measure such as $\log 1/C$, K_i , IC_{50} , ED_{50} , EC_{50} , $\log K$ and K_m .

The relationship or function is more often than not a mathematical expression derived by statistical or related techniques. In present study the multiple linear regression (MLR) technique is used. The parameters describing structural and physicochemical properties are used as independent variables and the biological activities are dependent variables.

In the present investigation a QSAR study is performed over a set of 21 phenyl acridine derivatives. This is the extension of previous studies based on the application of topological and physicochemical parameters in QSAR.²⁵⁻³⁴

Acridine derivatives are among the oldest classes of bioactive compounds, widely used as antibacterial and antiprotozoal agents. Some work in these areas continues, but recent research has mainly focused on their use as anticancer drugs, because of the ability of the acridine chromophore to intercalate into DNA and to inhibit topoisomerase enzymes.³⁵ These compounds also have been used as chemotherapeutic agents against cancer cells.³⁶ In the field of antitumor DNA-binding agents, this class of acridine derivatives play an important role both in the number of active compounds and in the importance of DNA binding affinity.³⁷

The DNA binding affinity ($\log K$) is then modeled using distance- based topological indices: Wiener (W)³⁸, Szeged (Sz)^{39,40} Balaban (J)⁴¹ indices and Randic connectivity index (χ)⁴² related to branching of molecules along with physicochemical properties: Molar Refractivity (MR), Molar Volume (MV), Parachor (Pc) etc. and indicator parameters accounted for the substitution effect on various positions.

The results obtained are better (discussed in the Result and Discussion part of this study) than those of a previous QSAR study performed by Hansch and coworkers⁴³ in their review work on applicability of hydrophobic parameters in QSAR.

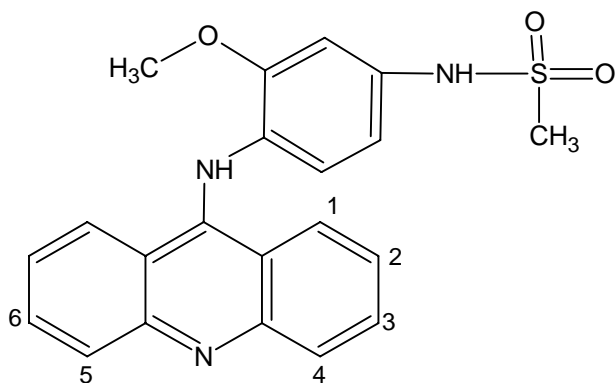


Figure 1. Parent structure of phenylamino-acridine derivatives used in present study.

Results and Discussion

The phenylamino-acridine derivatives (Figure 1), their DNA binding affinity and indicator parameters are given in Table 1 (see p. 141).

Various widely used topological indices tested in the present study are recorded in Table 2 (see p.142). Table 3 (see p. 143) contains the tested physicochemical properties. The details of calculation of these indices and the source of calculation of physicochemical properties are given in the experimental section of this paper.

In proposing QSAR models for modeling the DNA binding affinity we have used the maximum R^2 method.⁴⁴ Initially we have used Pogliani's quality factor $Q^{45,46}$ for investigating predictive power of the various parameters and finally we used the cross validation parameters to prove our findings. The comparisons of various models by cross validation parameters are shown in Table 5 (see p. 145)

The inter-correlation of various parameters and the correlation of parameters with biological activity ($\log K$) are presented in Table 4 (see p. 144) in a correlation matrix. As shown by the correlation matrix, DNA binding affinity has a good correlation with all four topological indices as well as with the physicochemical parameters like MR, MV, Pc. Correlation data of univariate correlation of Wiener index (W) ($r = 0.8118$), Randic connectivity index (χ) ($r = 0.832$), Balaban index (J) ($r = 0.84$), and Szged index (Sz) ($r = 0.824$) exhibits the significance of topological indices in QSAR study of Phenylamino-acridine derivatives. Also a good correlation with physicochemical properties like MR ($r = 0.886$), MV ($r = 0.862$) and Pc ($r = 0.869$) expresses the importance of the structural and volumetric features in theoretical modeling of DNA binding affinity ($\log K$).

This also shows the dependence of activity on structural features of the molecule as well as justifies the structural numerates of molecules in the form of topological indices.

For the QSAR study of the same series we tested the bivariate combinations of the parameters. The results obtained from the bivariate combinations are encouraging and better models are shown below with their statistics.

$$\log K = 2.2577 \times 10^{-4} (\pm 7.3634 \times 10^{-5}) W + 0.3477 (\pm 0.0942) I_{DS} + 5.528 \quad (1)$$

$n = 21, r = 0.9034, Se = 0.1509, Q = 5.99$

$$\log K = 0.1318 (\pm 0.0402) \chi + 0.3228 (\pm 0.0955) I_{DS} + 4.1899 \quad (2)$$

$n = 21, r = 0.9018, Se = 0.1474, Q = 6.16$

$$\log K = 5.070 (\pm 2.3533) J + 0.2876 (\pm 0.1433) I_{DS} - 1.3198 \quad (3)$$

$n = 21, r = 0.8818, Se = 0.166, Q = 5.31$

$$\log K = 1.4387 \times 10^{-4} (\pm 4.6158 \times 10^{-5}) Sz + 0.3309 (\pm 0.097) I_{DS} + 5.4601 \quad (4)$$

$n = 21, r = 0.9046, Se = 0.1501, Q = 6.03$

$$\log K = 0.0387 (\pm 0.004) MR + 0.2056 (\pm 0.0739) I_3 + 1.3645 \quad (5)$$

$n = 21, r = 0.928, Se = 0.1312, Q = 7.07$

$$\log K = 0.126 (\pm 0.0015) MV + 0.2341 (\pm 0.0813) I_3 + 2.1679 \quad (6)$$

$n = 21, r = 0.9112, Se = 0.145, Q = 6.28$

$$\log K = 0.0081 (\pm 0.0019) MV + 0.3076 (\pm 0.0802) I_{DS} + 3.6349 \quad (7)$$

$n = 21, r = 0.9293, Se = 0.13, Q = 7.15$

$$\log K = 0.0043 (\pm 4.5587 \times 10^{-4}) Pc + 0.2416 (\pm 0.0739) I_3 + 2.2095 \quad (8)$$

$n = 21, r = 0.927, Se = 0.132, Q = 7.02$

$$\log K = 0.0029 (\pm 6.455 \times 10^{-4}) Pc + 0.2874 (\pm 0.0823) I_{DS} + 3.6082 \quad (9)$$

$n = 21, r = 0.9308, Se = 0.1287, Q = 7.23$

$$\log K = 0.0267 (\pm 0.0053) MR + 0.2734 (\pm 0.0767) I_{DS} + 2.8931 \quad (10)$$

$n = 21, r = 0.9398, Se = 0.1203, Q = 7.81$

When earlier work⁴³ was repeated using an entirely new set of parameters, the new results should show either better statistics or be of an equivalent quality. In the present study the models (Eq 1 to 6) have competitive statistics with less and totally different parameters than the model proposed by Hansch and co-workers⁴³ in their review study on positive hydrophobic parameters.

The remaining biparametric models (Eq 7 to 10) containing combinations of MR, MV, PC, and indicator parameter I_{DS} and I_3 gives slightly better statistics with less standard error of estimation (Se) and higher Q value than that of model proposed by Hansch and co-workers⁴³ ($n = 21, r = 0.92951, Se = 0.134, Q = 6.94$). The models proposed in the present study are also better in terms of parameters because the present models have less parameter than the model proposed by Hansch and co-workers.⁴³

Eq 1 to 10 also exhibit the applicability of topological and physicochemical parameters for the QSAR study of phenylamino-acridine derivatives.

For detailed SAR studies we tested the trivariate combinations which resulted in an excellent improvement in statistics. The best triparametric model was obtained from the combination of MR, I_{DS} and I_3 as below.

$$\log K = 0.029 (\pm 0.0054) MR + 0.2085 (\pm 0.0861) I_{DS} + 0.1132 (\pm 0.0761) I_3 + 2.5501 \quad (11)$$

$n = 21, r = 0.947, Se = 0.1164, Q = 8.14$

The statistics obtained from Eq 11 demonstrates the role of the parameter MR in the modeling of DNA binding affinity. The parameter MR is a physicochemical parameter and is a combined effect of size and polarizability of the substituents. It characterizes deformation of

molecular electrons distribution. The equation also shows the direct relationship between MR and $\log K$, i.e., an increase in Molar Refractivity enhances the DNA binding of phenylamino-acridine derivatives. The equation also expresses the significance of the indicator parameter I_{DS} (which accounts for the presence of di-substitution) in DNA binding, i.e., di-substitution in phenylamino-acridine derivatives enhances their DNA binding affinity. Similarly the positive correlation coefficient of indicator parameter I_3 accounting for substitution at the 3rd position shows the direct relationship between substitution at the 3rd position and DNA binding affinity. The presence of the indicator parameter I_{DS} in most of the models and its high magnitude in eq 11 demonstrates the dominating role of di-substitution in DNA binding affinity of phenylamino-acridine derivatives as compared to other indicator parameter and topological parameters. The estimated DNA binding affinities from Eq. 11 are shown in Table 6 (see p. 146) and are graphically presented in Figure 2 (see p 147).

As opposed to traditional regression methods, the method of cross-validation estimates the trustworthiness of a model by predicting data. This method uses cross-validated fewer parameters: PRESS (predicted residual sum squares), SSY (sum of the squares of response value), r^2_{cv} (overall predictive ability), and adjusted r^2 .

PRESS (predicted residual sum of squares) is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Its value being less than SSY (sum of the squares of response value) points out that the model predicts better than chance and can be considered statistically important. In the present case all the proposed models have PRESS \ll SSY demonstrating them to be better than chance and statistically significant.

Furthermore, the ratio PRESS/SSY is used to estimate the confidence interval of the DNA binding affinity. To have a dependable QSAR model, PRESS/SSY should be smaller than 0.4. In our case the ratio PRESS/SSY ranges between 0.11-0.29 indicating that all the proposed models are reliable QSAR models.

The indication of the performance of the model is obtained from r^2_{cv} (the overall predictive ability). In our case, the highest r^2_{cv} is found for the model expressed by equation (11), indicating that it has an outstanding predictive power.

Conclusions

On the basis of the results discussed above, it can be concluded that the DNA binding affinity of phenylamino-acridine derivatives are structure specific in nature and most of the topological and structural parameters like W, $^1\chi$, Sz, MV, Pc and MR can be applicable for the modeling of DNA binding affinity of the phenylamino-acridine derivatives. It can be also concluded that the di-substitution and substitution at 3rd position have a significant role in the DNA binding affinity and the models proposed in the present investigation are better than the previously proposed models.⁴³

Experimental Section

Biological activity (logK)- DNA binding affinity expressed as logK, was taken from the literature.⁴³

Topological indices- All the topological indices used are calculated from the hydrogen suppressed molecular graphs. Though their calculations are exclusively discussed in the literature, we give below the expressions used for their calculations.

Wiener index (W)³⁸-Wiener index $W = W(G)$ of G is defined as the half sum of the elements of the distance matrix.

$$W = W(G) = \frac{1}{2} \sum_{i=1} \sum_{j=1} (D)_{ij} \quad (12)$$

Where, $(D)_{ij}$ is the ij th element of the distance matrix which denotes the shortest graph – theoretical distance between sites i and j of G .

The connectivity index (${}^1\chi$)⁴²- the connectivity index ${}^1\chi = {}^1\chi(G)$ of G is defined by Randić as

$${}^1\chi = {}^1\chi(G) = \sum_{i,j} [d(i) \cdot d(j)]^{-0.5} \quad (13)$$

Balaban index (J)⁴¹ - The Balaban index $J = J(G)$ of G is defined as

$$J = M/(\mu+1) \sum_{\text{Bonds}} (d_i \cdot d_j)^{-0.5} \quad (14)$$

Where M is the number of bonds in G , μ is the cyclomatic number of G , and d_i ($i = 1, 2, 3, N$; N is the number of vertices in G) is the distance sum.

The cyclomatic number $\mu = \mu(G)$ of a cyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related acyclic graph. In case of monocyclic graph $\mu = 1$ otherwise it is calculated by means of the following expression

$$M = M - N + 1 \quad (15)$$

Szeged index (Sz)^{39,40}- the Szeged index, $Sz = Sz(G)$, is calculated according to the following expression:

$$Sz = Sz(G) = \sum_{\text{Edges}} n_u \cdot n_v$$

Where n_u is the number of vertices lying closer to one end of the edge $e = uv$; the meaning of n_v is analogous. Edges equidistance from both the ends of an edges, $e = uv$ are not taken into account.

Physicochemical parameters- In present study molar refractivity (MR), molar volume (MV), parachor (P_c), index of refraction (η), surface tension (ST), density (D), shown in Table 3 are tested and calculated from the computer software acdlabs (Chem sketch 5.0).⁴⁶

Acknowledgements

Authors are thankful to Dr. PV Khadikar for his guidance to perform the study.

Authors are also thankful to Dr. C. Hansch, A. Kurup and R. Garg for providing the literature for the study.

References

1. Randić', M. *Chemometrics Intel. Lab. Systems* **1991**, 10, 213.
2. Randić', M. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 672.
3. Balaban, A. T.; Motoc, I.; Bonchev, D.; Mekenyan, O.; Charton, M.; Motoc, I. *Eds; Topics Curr. Chem.* **1983**, 114, 21.
4. Balaban, A. T. Historical developments of topological indices. *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds; Gordon and Breach: Amsterdam, The Netherlands, 1999; p 403.
5. Balaban, A. T. *J. Mol. Struct. (THEOCHEM)* **1988**, 165, 243.
6. Basak, S. C. Information theoretic indices of neighborhood complexity and their applications. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds; Gordon and Breach: Amsterdam, The Netherlands, 1999; p 563.
7. Balaban, A. T. From Chemical Graphs to 3D Molecular Modeling. In *From Chemical Topology to Three-dimensional Geometry*; Plenum Press: New York, 1977; p 1.
8. Trinajstić', N. In *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992; p 225.
9. Basak, S. C.; Grunwald, G. D.; Niemi, G. J. Use of graph-theoretic geometrical molecular descriptors in structure-activity relationships. In *From Chemical Topology to Three-dimensional Geometry*; Plenum Press: New York, 1977; p 73.
10. Randić', M.; Razinger, M. On characterization of 3D molecular structure. In *From Chemical Topology to Three-dimensional Geometry*; Plenum Press: New York, 1977; p 159.
11. Mihalić', Z.; Trinajstić', N. *J. Mol. Struct. (THEOCHEM)* **1991**, 232, 65.
12. Bogdanov, B.; Nikolić', S.; Trinajstić', N. On the three-dimensional Wiener number. *J. Math. Chem.* **1988**, 3, 299.
13. Bogdanov, B.; Nikolić', S.; Trinajstić', N. *J. Math. Chem.* **1990**, 5, 305.
14. Bogdanov, B. *J. Math. Chem.* **1989**, 3, 299.
15. Diudea, V.; Horvath, D.; Graovac, A. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 129.
16. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 708.

17. Zhu, H.-Y.; Klein, D. J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1067.
18. Randić, M.; Razinger, M. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 140.
19. Randić, M. *Stud. Phys. Theor. Chem.* **1988**, *54*, 101.
20. Randić, M. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1988**, *15*, 201.
21. Randić, M.; Jerman-Blazić, B.; Trinajstić, N. *Computer Chem.* **1990**, *14*, 237.
22. Randić, M. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 373.
23. Randić, M. *New J. Chem.* **1995**, *196*, 781.
24. Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1981**, *70*, 583.
25. Thakur, A.; Thakur, M.; Khadikar, P. *Bioorg. Med. Chem.* **2003**, *11*, 5203.
26. Thakur, A.; Thakur, M.; Khadikar, P.; Supuran, C. T.; Sudele, P. *Bioorg. Med. Chem.* **2004**, *12*, *4*, 789.
27. Thakur, M.; Thakur, A.; Khadikar, P. *Bioorg. Med. Chem.* **2004**, *12*, *4*, 825.
28. Thakur, A.; Thakur, M.; Vishwakarma, S. *Bioorg. Med. Chem.* **2004**, *12*, *5*, 1209.
29. Thakur, A.; Thakur, M. *Bioinformatics India* **2003**, *1*, 63.
30. Thakur, M.; Thakur, A.; Sudele, P. *Indian Journal of Chemistry sec B*, in press.
31. Thakur, A.; Thakur, M. *Res. J. Chem. Environ.* **2003**, *7*, *1*, 51.
32. Thakur, A.; Agrawal, A.; Thakur, M.; Sudele, P. *Bioinformatics India* **2003**, *1*, *3*, 18.
33. Thakur, A. *Ult. Scie. Phy. Scie.* **2003**, *15*, *3*, 407.
34. Thakur, M.; Agrawal, A.; Thakur, A.; Khadikar, P. *Bioorganic and Medicinal Chemistry* **2003**, *12*, *9*, 2287.
35. William, A. D. *Current Medicinal Chemistry* **2002**, *9*, *18*, 1655.
36. Adrienne, A. *Current Medicinal Chemistry* **2002**, *9*, *18*, 1667.
37. Ippolito, A. *Current Medicinal Chemistry* **2002**, *9*, *18*, 1701.
38. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
39. Gutman, I. *Graph Theory Notes, New York* **1994**, *27*, 9.
40. Khadikar, P. V.; Deshpande, N. N.; Kali, P. A.; Dohirnin, A.; Gutman, I.; Donewfor, G., *J. Chem. Inf. Compt. Sci.* **1995**, *35*, 547.
41. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
42. Randić, M. *Croat. Chem. Acta.* **1993**, *66*, 289.
43. Hansch, C.; Kurup, A.; Garg, R. *Hua Gao Chem. Rev.* **2001**, *101*, 619.
44. Pogliani, L. *Amino Acid* **1994**, *6*, 141.
45. Pogliani, L. *J. Pharm. Chem* **1996**, *100*, 18065.
46. Computer software ChemsSketch 5.0 (from ACD labs)