Bioorganic & Medicinal Chemistry Letters 19 (2009) 1737–1739

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Comparative QSAR modelling of 2-phenylindole-3-carbaldehyde derivatives as potential antimitotic agents

Amit Kumar Halder, Nilanjan Adhikari, Tarun Jha *

Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, P.O. Box 17020, Jadavpur University, Kolkata, West Bengal 700 032, India

ARTICLE INFO

Article history:
Received 7 July 2008
Revised 22 January 2009
Accepted 26 January 2009
Available online 29 January 2009

Keywords:
Antimitotic agents
2-Phenylindole-3-carbaldehyde derivatives
Electrostatic potential charges
MLR-FA
QSAR modelling

ABSTRACT

QSAR modelling was done on some 2-phenylindole-3-carbaldehyde derivatives to find out structural requirements for more active antimitotic agents. Four statistical methods were used to develop models. The results show the importance of ETSA indices, RTSA indices, *IC1*, *SIC4*, *Jhetv* and *MSD* on the activity. Electrostatic potential charges of atoms, increased surface area, and presence of bulky group along *Y-axis* and chlorine substitution were also found to be important.

Mitosis is a part of cell cycle where chromosomes are divided into separate parts of cells.[1,2] Microtubules are responsible for capturing and aligning of chromosomes in metaphase and separating to daughter cells in anaphase.[3] Microtubules are targets for antimitotic agents. Antimitotic agents like vinca alkaloids, taxanes, etc., are used for the treatment of cancers. Untoward adverse effects, difficulty in synthesis and high cost limited their uses.[4] The search for new antimitotic agents with better activity is still in progress. To find out the structural requirements for more active antimitotic agents, QSAR study was performed on some 2-phenylindole-3-carbaldehyde derivatives. The general structure of these compounds is shown in Figure 1.

50% inhibitory concentration [IC_{50}] of 33 substituted 2-phenylindole-3-carbaldehydes on MDA-MB231 breast cancer cells was collected.[5] IC_{50} values were converted to the negative logarithmic scale [pIC_{50}]. Anti-proloferative activity data (pIC_{50}) are given in Table 1. Electrotopological state atom (ETSA) index[6,7] and refractotopological state atom (RTSA) index[8] were calculated by using the program 'Mouse'.[9] *Hyperchem. Release 7.0 Pro Package*[10] was used for the calculation of molar volume (*V*), approximate surface area (*SAA*), surface area grid (*SAG*), hydrophobicity (*logP*), molecular polarizability (*MP*), molecular mass (*MASS*) and molecular orbital energies (HOMO and LUMO). The energy minimizations were done by using molecular mechanical (MM^+) force fields without cut-off for non-bonded interactions, solvation and constrains. These were geometrically optimized by AM1 (Austin model 1) method. *Dragon*[11] was utilized for the calculation of other descriptors. Descriptors[12–20] showing significant correlation with biological activity were chosen for QSAR study. Electrostatic potential (*EP*) charges of each atom and principal moment of inertia towards *X*-, *Y*- and *Z*-*axes* were calculated using *Chem. 3D Pro* package.[21] Electrostatic potential charges[22] were calculated by AM1 method whereas energy minimization of was done by RHF (Restricted Hartree–Fock: closed shell) wave function.

Statistical qualities of equations were justified by correlation coefficient R, adjusted R^2 (R_a^2), variance ratio (*F*), probability factor related to *F*-ratio (*p*) and standard error of estimate (*s*). Leave-One-Out (LOO) cross-validation method was used to validate the predictive powers of all QSAR equations. The predicted residual sum of square (PRESS), cross validated R^2 (R_{cv}^2), standard deviation of PRESS (SDEP) and standard deviation error of prediction (S_{PRESS}) were considered for the validation of QSAR models.

Principal component regression analysis (PCRA)[23]: Nine factor scores were extracted and rotated by VARIMAX rotation. These fac-

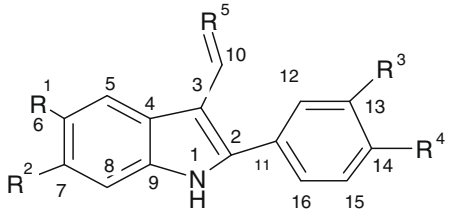

Figure 1. The general structure of 2-phenylindole-3-carbaldehyde compounds.

* Corresponding author. Tel.: +91 33 24146666x2495; fax: +91 33 24146927.
E-mail address: tjupharm@yahoo.com (T. Jha).

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.01.081

Table 1
Anti-proliferative data for substituted 2-phenyl indole-3-carbaldehydes

Compound	R^1	R^2	R^3	R^4	R^5	IC_{50}	pIC_{50}
1	OMe	H	H	OMe	O	260	6.586
2	H	OMe	H	OMe	O	35	7.456
3	H	F	H	OMe	O	59	7.229
4	F	H	H	OMe	O	540	6.268
5	H	F	H	OMe	O	27	7.569
6	Me	Cl	H	OMe	O	26	7.585
7	Me	H	H	OMe	O	86	7.065
8	*n*-Pr	H	H	OMe	O	20	7.699
9	*i*-Pr	H	H	OMe	O	29	7.537
10	*n*-But	H	H	OMe	O	6.7	8.174
11	*i*-But	H	H	OMe	O	72	7.143
12	*t*-But	H	H	OMe	O	280	6.553
13	*n*-Pent	H	H	OMe	O	5.5	8.260
14	*n*-Hex	H	H	OMe	O	7.4	8.131
15	H	OMe	OMe	H	O	1030	5.987
16	H	OMe	OMe	OMe	O	270	6.569
17	H	OMe	OH	OMe	O	800	6.097
18	H	OMe	H	Me	O	31	7.509
19	H	Cl	H	Me	O	7.8	8.108
20	Me	H	H	Me	O	48	7.319
21	*n*-But	H	H	Me	O	34	7.468
22	*n*-But	H	H	Et	O	27	7.569
23	Et	H	H	n-But	O	300	6.523
24	*n*-But	H	H	F	O	350	6.456
25	*n*-But	H	H	CF_3	O	33	7.481
26	*n*-Pent	H	H	CF_3	O	42	7.377
27	*n*-Hex	H	H	CF_3	O	43	7.366
28	H	OMe	H	OMe	NCH_3	34	7.468
29	*n*-But	H	H	OMe	NCH_3	6	8.222
30	*n*-Pent	H	H	OMe	NCH_3	6	8.222
31	*n*-But	H	H	CF_3	NCH_3	32	7.495
32	*n*-But	H	H	OMe	N-OH	40	7.398
33	*n*-But	H	H	CF_3	N-OH	497	6.304

tor scores were used as the independent parameters. The following equation was developed:

$$pIC_{50} = 7.278(\pm0.062) + 0.208(\pm0.063)f_1 - 0.221(\pm0.063)f_2 + 0.370(\pm0.063)f_6 - 0.168(\pm0.063)f_7 - 0.274(\pm0.063)f_8 \quad (1)$$

$n = 33$; $R = 0.864$; $R^2 = 0.746$; $R^2_A = 0.700$; $F(5,27) = 15.905$; $p < 0.00001$; S.E.E = 0.357; PRESS = 5.115; $R^2_{cv} = 0.624$; SDEP = 0.394; $S_{PRESS} = 0.435$.

Where n is the number of data points. Eq. (1) explains 74.0% variance and predicts 67.5% variance. The 95% confidence intervals of regression coefficients are shown in parentheses. Eq. (1) shows the importance of factors 1, 2, 6, 7 and 8. Factor 1 is highly loaded with *SAA*, *SAG*, *Vol*, *logP*, *Refract*, *Polariz*, *PMY*, *PMZ*, *CIC1*, *IC4*, *IC5*, *MW*, *nSK*, *nBT*, *nBO*, *SCBO*, *RBN*, *RBF*, *nH* and *nC*. It shows importance of these descriptors. Factor 2 shows importance of S_2, S_{11}, S_{12}, E_{HOMO}, R_{11}, R_{12}, R_{15}, R_{16}, *IC0*, nF and *nX*. Factor 6 is highly loaded with R_{12}, EP_{11}, EP_{13}, EP_{15} and EP_{16} whereas factor 7 shows significance of IC_2. Factor 8 shows importance of I_1 and *nCl*.

Partial least square (PLS)[24]: The number of latent variables was two by cross-validation method. The predicted variables were selected on the basis of regression coefficients. Variables with smaller standardized coefficients were removed from PLS regression. PLS models were developed using larger standardized coefficients. The equation obtained was:

$$pIC_{50} = 4.334 + 0.390S_6 + 0.467I_1 - 1.277EP_4 + 1.455EP_8 - 2.1204EP_{11} + 2.727EP_{12} - 2.357EP_{15} + 2.175EP_{16} + 25.231MSD - 0.948IC1 \quad (2)$$

$n = 33$; $R = 0.891$; $R^2 = 0.794$; $R^2_A = 0.686$; $F = 57.94$; $p < 0.00001$; S.E.E = 0.357; PRESS = 4.402; $R^2_{cv} = 0.600$; SDEP = 0.365; $S_{PRESS} = 0.471$.

Eq. (2) explains 68.6% variance and predicts 60% of variances of biological activity. S_6 is *E*-state index of the atom number 6. The positive coefficient indicates that the higher value may be conducive to the activity. Indicator parameter I_1 stands for the presence or absence of chlorine atom at R^2 position. The positive coefficient of I_1 signifies chlorine atom at R^2 position may be favorable for antimitotic activity. EP_4, EP_8, EP_{11}, EP_{12}, EP_{15} and EP_{16} are electrostatic potential charges of atom numbers 4, 8, 11, 12, 15 and 16, respectively. Positive coefficients of EP_8, EP_{12} and EP_{16} indicate that the higher negative electrostatic charges at atom numbers 8, 12 and 16 may be conducive to the biological activity. Negative coefficients of EP_4 and EP_{15} suggest that the higher electrostatic potential charges of these atom numbers 4 and 15 may be detrimental to the affinity. The negative electrostatic potential of the atom number 11 indicates that this atom may be more susceptible to electrophilic attack. *IC1* stands for information content index (neighborhood symmetry of order 1).[15] The negative coefficient suggests that the lower value may correspond to the higher activity. Eq. (2) also shows the importance of *MSD*[19] (Balaban-type mean square distance index). The positive coefficient indicates that the higher value may be favorable for the affinity.

Stepwise regression[24]: Using stepping criteria on the basis of *F* value ($F = 3.0$ for inclusion; $F = 2.9$ for exclusion), the following equation was derived:

$$pIC_{50} = 6.252(\pm2.086) - 1.550(\pm0.258)E_{HOMO} + 0.002(\pm0.001)SAA - 0.517(\pm0.057)R_7 + 13.844(\pm1.164)EP_{16} - 2.780(\pm0.292)IC1 - 1.395(\pm0.321)SIC4 \quad (3)$$

$n = 33$; $R = 0.951$; $R^2 = 0.904$; $R^2_A = 0.881$; $F(6,26) = 40.692$; $p < 0.00001$; S.E.E = 0.224; PRESS = 2.130, $R^2_{cv} = 0.843$; SDEP = 0.254; $S_{PRESS} = 0.286$.

R_7 and *SIC4* stand for *R*-state index of the atom number 7 and structural information content index (neighborhood symmetry of order 4),[17] respectively. E_{HOMO} and *SAA* are the energy of the highest occupied molecular orbital and approximate surface area of the molecule, respectively. Eq. (3) predicts 84.3% variance and explains 88.1% of variance of the affinity. The negative coefficient of E_{HOMO} indicates that the affinity may increase with the decrease in the electron donating property. The positive coefficient of *SAA* suggests that the higher activity may be due to the higher molecular surface area. The negative coefficient of R_7 indicates that the lower value of *R*-state index of the atom number 7 may improve the activity. The negative coefficient of *SIC4* suggests that the affinity may increase with the decrease of *SIC4*.

Factor analysis-multiple linear regression (FA-MLR)[23,24]: Factor analysis was performed as preprocessing step to select descriptors for QSAR equations. Nine factors could explain the data matrix to the extent of 93.20%. pIC_{50} was highly loaded with factor 6 (highly loaded with EP_{13} and EP_{15}), and factor 8 (highly loaded with I_1 and *nCl*), moderately loaded with factor 2 (highly loaded with S_2, S_{11}, S_{12}, S_{15}, S_{16}, R_{11}, R_{15}, R_{16}, *IC0*, *Ss*, *Me*, *Ms* and *nF*), factor 1 (highly loaded with *SAA*, *SAG*, *Vol*, *logP*, *Refr*, *Polar*, *MASS*, *PCIY*, *PCIZ*, *CIC1*, *IC4*, *IC5*, *Mw*, *Mv*, *Mp*, *nAT*, *nBT*, *RBN*, *RBF*, *nH* and *nC*), factor 7 (considerably loaded with *IC2*), factor 4 (highly loaded with S_{10}, R_{10}, EP_2, EP_3 and *nN*) and poorly loaded with factor 3, factor 5 and factor 9. Different combinations of parameters having factor loading of more than 0.6 were subjected to multiple regression. Inter-correlated parameters were not considered for the development of equations. The best pentavariate FA-MLR equation is:

$$pIC_{50} = 46.165(\pm5.229) - 8.034(\pm0.916)R_4 + 3.630(\pm0.852)EP_3 - 5.014(\pm1.008)EP_{15} - 7.715(\pm1.312)Jhetv - 21.283(\pm3.417)RBF \quad (4)$$

$n = 33$; $R = 0.914$; $R^2 = 0.836$; $R^2_A = 0.806$; $F(5,27) = 27.556$; $p < 0.00001$; S.E.E = 0.287; $R^2_{cv} = 0.757$; PRESS = 3.307; SDEP = 0.316; $S_{PRESS} = 0.349$.

Eq. (4) predicts 75.7% variance and explains 80.6% of variances of the activity. R_4 is *R*-state index of the atom number 4. The negative coefficient of R_4 indicates that the increase in the value may be unfavorable to the activity. Eq. (4) shows the significance of *Jhetv*, i.e., a Balaban-type index from van der Waals weighted distance matrix. The negative coefficient indicates that the increase in van der Waals weighted distance may be detrimental to the affinity. EP_3 is the electrostatic potential charge at the atom number 3 and the positive coefficient suggests that the increase in the electrostatic charge of this atom may improve the activity. *RBF* corresponds to the rotatable bond fraction[20] and the negative coefficient indicates that fractional increase in the rotatable bonds may be unfavorable for the activity.

FA-MLR approach gave rise to another model:

$$\begin{aligned} \mathrm{pIC}_{50} = {} & 16.516(\pm 0.928) + 1.509(\pm 0.205)I_1 \\ & - 4.748(\pm 1.074)EP_4 + 11.500(\pm 1.190)EP_{16} \\ & - 2.560(\pm 0.316)IC1 - 1.293(\pm 0.371)SIC4 \\ & + 0.125(\pm 0.047)nF \end{aligned} \quad (5)$$

$n = 33$; $R = 0.931$; $R^2 = 0.867$; $R^2_A = 0.837$; $F(6,26) = 28.385$; $p < 0.00001$; S.E.E = 0.263; PRESS = 2.860; $R^2_{cv} = 0.790$; SDEP = 0.294; $S_{PRESS} = 0.332$.

Eq. (5) explains 83.7% variance and predicts 79.0% of variances of the activity. *nF* stands for the number of fluorine atom. The positive coefficient indicates that the increase in the number may improve the activity.

However, the best equation obtained is:

$$\begin{aligned} \mathrm{pIC}_{50} = {} & 14.795(\pm 0.891) - 0.441(\pm 0.052)R_7 \\ & + 0.450(\pm 0.115)R_{14} + 0.0001(\pm 0.000)PMY \\ & - 11.861(\pm 1.057)EP_{15} - 3.761(\pm 0.361)IC1 \\ & - 1.997(\pm 0.312)SIC4 \end{aligned} \quad (6)$$

$n = 33$; $R = 0.953$; $R^2 = 0.908$; $R^2_A = 0.887$; $F(6,26) = 42.967$; $p < 0.00001$; S.E.E = 0.219; PRESS = 1.819; $R^2_{cv} = 0.866$; SDEP = 0.235; $S_{PRESS} = 0.264$.

Eq. (6) predicts 86.6% variance and explains 88.7% of variance of the biological activity. Eq. (6) reveals the significance of R_7, R_{14}, *PMY* and EP_{15}. R_{14} is *R*-state index of the atom number 14 and *PMY* signifies the principle moment of inertia along *Y-axis*. The positive coefficient of R_{14} indicates that the increase in the value of *R*-state index of the atom number 14 may be conducive to the activity. The positive coefficient of *PMY* suggests that the presence of more bulky group towards *Y-axis* may correspond to the higher activity. All regression coefficients of Eqs. (1)–(6) are significant at more than 95% confidence level as supported by their *p*- and *t*-values.

In summary, this study explored the importance of topological, constitutional, geometrical and electronic parameters towards antimitotic activities of 2-phenylindole-3-carbaldehydes. Four methods were used that showed comparable results with satisfactory statistical quality. The study shows that the presence of chlorine atom at R^2 position and the presence of bulky group along *Y-axis* may be favorable for the affinity. Increase in the surface area and decrease in the rotatable bond fraction may be conducive to activity. QSAR models as also shown the negative contribution of R_7 as also shown by the SAR of the original work[5] as well as those

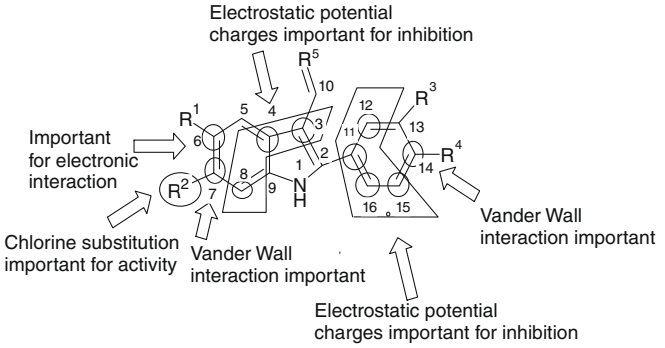

Figure 2. Important atoms and substituents of 2-phenyl-indole-3-carbaldehyde for antimitotic activity.

of *IC1* and *SIC4* towards the activity. The negative coefficient of E_{HOMO} may indicate the higher activity. The study also revealed the significance of electrostatic potentials of atom numbers 3, 4, 8, 11, 12, 15 and 16. Atoms and substituents important for antimitotic activity are shown in Figure 2.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2009.01.081.

References and notes

1. Lodish, H.; Berk, A.; Matsudaira, P.; Kaiser, C. A.; Krieger, M.; Scott, M. P.; Zipurski, S. L.; Darnell, J.. In *Molecular Cell Biology*; W.H. Freeman: New York, 2004. Chapter 20.
2. Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D.. In *Molecular Biology of the Cell*; Garland: New York, 1994. Chapter 18.
3. Jordan, M. A.; Wilson, L. *Nat. Rev. Cancer* **1987**, *4*, 253.
4. Tripathi, K. D. *Essentials of Medical Pharmacology*; Jaypee: New Delhi, 2003. Chapter 60.
5. Kaufmann, D.; Pojarova, M.; Vogel, S.; Liebl, R.; Gastpar, R.; Gross, D.; Nishino, T.; Pfaller, T.; Angerer, E. *Bioorg. Med. Chem.* **2007**, *15*, 5122.
6. Hall, H.; Mohney, B.; Kier, L. B. *Quant. Struct. – Act. Relat.* **1991**, *10*, 43.
7. de Gregorio, C.; Kier, L. B.; Hall, L. H. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 557.
8. Carrasco, R.; Padron, A. J.; Galvez, J. *J. Pharm. Pharmaceut. Sci.* **2004**, *7*, 19.
9. *Mouse*: a computer program developed by Department of Pharmaceutical Technology of Jadavpur University, India.
10. Hypercube, Inc. *Hyperchem*, Professional Release 7.0 Hypercube Inc., Gainesville.
11. *DRAGON* web version 2.1 is developed by Milano Chemometrics and QSAR Research Group, Dipartimento di Scienze dell'Ambiente e del Territorio Universitàdegli Studi di Milano, Bicocca.
12. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
13. Barysz, M.; Jashari, G.; Lall, R. S.; Srivastava, A. K.; Trinajstic, N. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; p 222.
14. Randic, M. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 607.
15. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures*; RSP-Wiley: Chichetser, 1983.
16. Magnuson, V. R.; Harriss, D. K.; Basak, S. C. In *Studies in Physical and Theoretical Chemistry*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; p 178.
17. Basak, S. C.; Gute, G. D. *SAR QSAR Environ. Res.* **1997**, *7*, 1.
18. Lovasz, L.; Pelikan, J. *Period. Math. Hung.* **1973**, *3*, 175.
19. Balaban, A. T. *Pure Appl.Chem.* **1983**, *55*, 199.
20. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley–VCH: Weinheim, 2000.
21. *Chem 3D* Pro Version 5.0 and *Chem Draw* Ultra Version 5.0, Programs of CambridgeSoft Corporation, USA.
22. Leach, A. R. *Molecular Modeling: Principles and Applications*; Pearson: London, 2001. Chapter 4.
23. Franke, R. *The Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.
24. (a) Leonard, J. T.; Roy, K. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4467; (b) Leonard, J. T.; Roy, K. *Bioorg. Med. Chem.* **2006**, *14*, 1039.