Asian Journal of Chemistry Vol. 22, No. 6 (2010), 4675-4680

2D-Quantitative Structure Property Relationship Study of N-(Aryl)-2-thiophene-2-ylacetamide Derivatives as Antitubercular Agents

H.S. CHOUHAN*, S.K. BHARTI and S.K. SINGH
Department of Pharmaceutics, Institute of Technology, Banaras Hindu University, Varanasi-221 005, India
Fax: (91)(542)2368428; Tel: (91)(542)6702749
E-mail: hschouhan.rs.phe@itbhu.ac.in

A series of N-(aryl)-2-thiophene-2-ylacetamide derivatives reported to possess antitubercular activity, have been taken for quantitative structure property relationship (QSPR) study. Quantitative structure property relationship of the series was studied between descriptors representing the molecular structures and partition coefficient by the stepwise linear regression analysis using OpenStat software version 7.07.09 and results were cross validated by Valstat software for its significance. The statistically significant relationships obtained are as follows: -log P = 0.689 (± 0.111) RsHa - 0.533 (± 0.231) Rs*f* - 2.3983. n = 21, r = 0.801, r^2 = 0.641155, var. = 0.0853691, σ = 0.29218, SE = 0.292, F = 16.0805, q^2 = 0.510649, r^{2bs} = 0.660584, S_{PRESS} = 0.341199, S_{DEP} = 0.315889. The result reveals that the hydrogen acceptor group substitutions on phenyl ring have favorable effect while field effects of substituents have detrimental effect on the partition coefficient of the molecule.

Key Words: 2D-Quantitative structure property relationship, Antitubercular agent, Lipophilicity.

INTRODUCTION

Tuberculosis (TB) is one of the most prevalent infectious diseases causing major health problem to human being. About 32 % of the world's population is infected with TB. Every year, approximately 8 million of these infected people develop active TB. It is the leading cause of death among the infectious diseases with about 2.0 million deaths every year worldwide[1]. It is caused by *Mycobacterium tuberculosis*. There are a number of effective drugs available for treating tuberculosis (TB) but current treatments are greatly complicated by the several months of chemotherapy required to eliminate persistent bacteria. In addition, widespread non-compliance has contributed to the emergence of multidrug-resistant (MDR) TB and extremely drug resistant (XDR) TB[2]. In the recent past, strategies to treat/manage TB have receiving much attention in terms of searching new lead or improving the effectiveness of the existing drugs. Scientists have begun investigating new drugs to control and eventually eliminate this scourge. Until now, progress in TB drug development has been impeded by several factors including MDR, long-term regimens, costs

and side effects associated with them. There is now recognition that new drugs to treat TB are urgently required, specifically for use in shorter treatment regimens than are possible with the current agents and which can be employed to treat multi-drug-resistant and latent disease.

Now, it is well accepted that, besides being pharmacologically active, an ideal drug should have some features regarding its bioavailability and its toxicological profile. *In silico* ADME/Tox filters are nowadays widely used to determine whether the new drug candidate is suitable to reach its site of action or to elicit toxic effects at its therapeutic dose. Moreover, modern approaches developed for a rational molecular design have moved the ADME/Tox *in silico* valuations to the early stages of drug development, where an optimal activity of the compound is searched[3].

It is known that the degree of absorption of a substance depends simultaneously on dose, solubility and permeability. The widely used Lipinski's 'rule of five' for orally bioavailable drugs describes the absorption by passive diffusion through the gastrointestinal barrier[4]. These simple rules state that oral bio-availability is likely to occur if at least three of the following rules are obeyed: molecular weight below 500; no more than five hydrogen bond donors and less than 10 hydrogen bond acceptors and calculated octanol-water partition coefficient (c log P) below 5. The conditions to satisfy Lipinski's rule and to manifest a good oral bioavailability involve a balance between the solubility of a compound and its ability to diffuse passively through the different biological barriers.

The solubility governs both the rate of dissolution of the compound and the maximum concentration reached in the gastrointestinal fluid. However, excessively polar compounds would result problematic at the stage of passing through the various biological barriers. Furthermore, many other reasons make solubility an important parameter in medicinal chemistry: soluble compounds are associated to shorter metabolization and elimination times, thus leading to lower toxicity and side effects. The most preclinic tests involve solubilization of the drug being tested in hydrophilic solvents[5,6]. Therefore, here a theoretical model such as quantitative structure-property relationships (QSPR) has been proposed to predict the best suited drug candidate for TB.

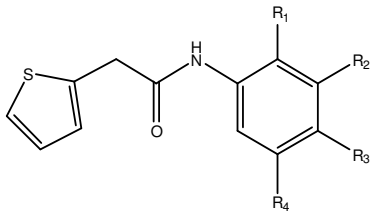
In the present study, a series of N-(aryl)-2-thiophen-2-ylacetamide derivatives (Table-1) which is already synthesized and evaluated for their antitubercular activity, has been taken for QSPR study[7]. These derivatives are selectively targeted to *M. tuberculosis* growth and are not cytotoxic to host cells at the concentrations effective in inhibiting *M. tuberculosis* infection. In this context, QSPR of the series was studied between descriptors representing the molecular structures and partition coefficient by the stepwise linear regression analysis.

EXPERIMENTAL

All calculations were run on a Pentium-IV personal computer with windows XP operating system. A stepwise multiple linear regression (MLR) procedure was

used for selection of descriptors using the OpenStat software version 7.07.09. The QSPR model for the calculation of the log P of a series of N-(aryl)-2-thiophen-2-ylacetamide derivatives is developed in a number of steps. The partition coefficient data for all 21 compounds has been taken from literature[7] and it was converted into -log P values. The descriptors representing molecular property were taken from the literature[8]. The structural descriptors were arranged in the excel sheet and saved as coma delimited file. The structural descriptors were selected and taken as independent variable and the -log P values for a set of all 21 compounds as dependent variables. Structure -log P model was generated by the multiple linear regression and statistical analysis using OpenStat software version 7.07.09[9]. The significant equation obtained after multiple linear regression analysis was validated by Valstat software for its significance.

TABLE-1
STRUCTURE REPRESENTING N-(ARYL)-2-THIOPHEN-2-YLACETAMIDE DERIVATIVES



Compound	R_1	R_2	R_3	R_4	R_5	log P
1	H	H	H	H	H	2.55
2	CH_3	H	H	H	H	2.74
3	OCH_3	H	H	H	H	2.66
4	NO_2	H	H	H	H	2.81
5	CF_3	H	H	H	H	3.86
6	H	Cl	H	H	H	3.55
7	H	OCH_3	H	H	H	2.71
8	H	CF_3	H	H	H	3.94
9	H	NO_2	H	H	H	2.86
10	H	H	CH_3	H	H	2.96
11	H	H	F	H	H	3.01
12	H	H	Br	H	H	3.31
13	H	H	OCH_3	H	H	2.75
14	H	H	Cl	H	H	3.44
15	H	H	NO_2	H	H	2.77
16	CH_3	H	H	H	CH_3	2.92
17	Cl	H	H	H	Cl	3.96
18	Cl	H	F	H	H	3.56
19	F	H	Cl	H	H	3.41
20	OCH_3	H	H	OCH_3	H	2.67
21	H	OCH_2O	H	H	H	2.66

RESULTS AND DISCUSSION

The number of descriptors in the final QSPR model were determined on the basis of the data set size (n = 21 compounds, Table-1), correlation coefficient, F-values and standard deviation obtained. In this way, the best two-parameter correlation was obtained for the entire data set of N-(aryl)-2-thiophen-2-ylacetamide derivatives. The best model is selected from the various statistically significant equations on the basis of the observed squared correlation coefficient (r^2), the standard error of the estimate (SE), sequential Fischer test (F), the bootstrapping squared correlation coefficient (r^{2bs}), the cross-validated squared correlation coefficient using leave-one-out (LOO) procedure (q^2), chance statistics (evaluated as the ratio of the equivalent regression equations to the total number of randomized sets; a chance value of 0.001 corresponds to 0.1 % chance of fortuitous correlation), outliers (on the basis of Z-score value) and the squared correlation coefficient of test set (r^2), standard predicted residual sum of squares (S_{PRESS}), standard deviation of error of prediction (S_{DEP}). The statistically significant equation was obtained and as follows:

$$-\log P = 0.689\ (\pm\ 0.111)\ RsHa - 0.533\ (\pm\ 0.231)\ Rsf - 2.3983$$

where n = 21, r = 0.801, r^2 = 0.641155, var. = 0.0853691, σ = 0.29218, SE = 0.292, F = 16.0805, q^2 = 0.510649, r^{2bs} = 0.660584, S_{PRESS} = 0.341199, S_{DEP} = 0.315889.

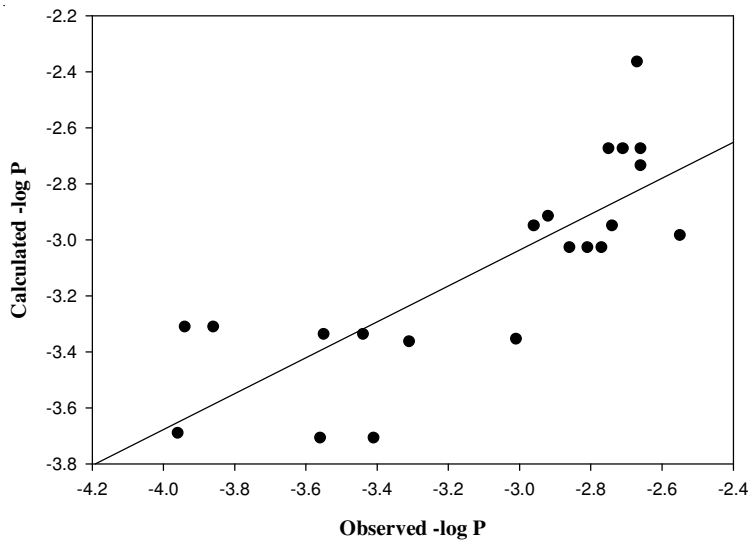


Fig. 1. Plot of experimental *versus* calculated value of -log P

In order to select robustness and the practical applicability of significant QSPR equations, a number of statistical approaches utilized along with the determination of high correlation coefficient. Here, internal consistency of the training set was confirmed by using LOO cross-validation method to ensure the robustness of the equations.

In the above significant equation having a correlation (r = 0.801) between partition coefficient with hydrogen acceptor (Ha) and field value (*f*). The developed QSPR model has a better correlation coefficient (r = 0.801), which accounts for more than 64.1 % of the variance in the activity. Also the inter-correlation among the parameters being less (< 0.160). No compound was found to be outliers when the equation tested for outlier by the Z-score method (Table-2), indicating that the equation is sufficient to explain the structurally diverse analogs of the series and is helpful in designing more potent compounds using physicochemical parameters. The LOO cross-validation method was employed for the prediction of the -log P and q^2 value (in the partition coefficient data of leave one compound) of 0.510 corresponds to a confidence limit greater than 95 %, which minimizes the risk of finding a significant explanatory equation for the partition coefficient just by mere chance. The predictive residual sum of square (S_{PRESS} = 0.341) and standard error of prediction (S_{DEP} = 0.315889) suggested a good internal consistency as well as predictive ability of the partition coefficient with low S_{DEP}. The r^{2bs} with the conventional squared correlation coefficient (r^2) and randomized partition coefficient test (chance < 0.001) revealed that the result was not based on chance correlation. The robustness and wide applicability of the model were further explained by a significant r^2 of the test set data. Hence, it is very clear that above equation fulfills the statistical validation criteria to a significant extent and to be a useful theoretical basis for proposing compounds with better lipophilicity.

TABLE-2
OBSERVED, CALCULATED AND PREDICTED -log P
(BY THE LEAVE ONE OUT PROCEDURE (LOO) METHOD) OF
N-(ARYL)-2-THIOPHEN-2-YLACETAMIDE DERIVATIVES

Compound	-log P	Calculated value	Residual	Z- value	Predicted -log P (LOO)	Residual
1	-2.55	-2.98330	0.43330	1.6018300	-3.06478	0.51478
2	-2.74	-2.94889	0.20889	0.7722300	-2.99404	0.25404
3	-2.66	-2.67348	0.01348	0.0498240	-2.67523	0.01523
4	-2.81	-3.02619	0.21619	0.7992020	-3.05705	0.24705
5	-3.86	-3.31021	-0.54979	-2.0324600	-3.26665	-0.59335
6	-3.55	-3.33602	-0.21398	-0.7910530	-3.31881	-0.23119
7	-2.71	-2.67348	-0.03652	-0.1350150	-2.66874	-0.04126
8	-3.94	-3.31021	-0.62979	-2.3282000	-3.26031	-0.67969
9	-2.86	-3.02619	0.16619	0.6143630	-3.04991	0.18991
10	-2.96	-2.94889	-0.01111	-0.0410600	-2.94649	-0.01351
11	-3.01	-3.35322	0.34322	1.2688100	-3.38133	0.37133
12	-3.31	-3.36182	0.05182	0.1915790	-3.36612	0.05612
13	-2.75	-2.67348	-0.07652	-0.2828860	-2.66356	-0.08644
14	-3.44	-3.33602	-0.10398	-0.3844080	-3.32766	-0.11234
15	-2.77	-3.02619	0.25619	0.9470730	-3.06276	0.29276
16	-2.92	-2.91448	-0.00552	-0.0203982	-2.91311	-0.00689
17	-3.96	-3.68873	-0.27127	-1.0028400	-3.62060	-0.33940
18	-3.56	-3.70593	0.14593	0.5394780	-3.74525	0.18525
19	-3.41	-3.70593	0.29593	1.0939900	-3.78567	0.37567
20	-2.67	-2.36365	-0.30635	-1.1325000	-2.16503	-0.50497
21	-2.66	-2.73370	0.07370	0.2724400	-2.74203	0.08203

Further, the obtained equation reveals that the substitution by hydrogen acceptor on all positions of the phenyl ring (RsHa) in N-(aryl)-2-thiophen-2-ylacetamide derivatives may have positive contribution to the lipophilicity of the compounds while field effect (Rs*f*) of substituent have detrimental effect on the partition coefficient of the molecule.

Conclusion

In the present work, a good 2D-QSPR model was developed for the prediction of partition coefficient with significant structural differences. On the basis of acceptable correlation (r = 0.801), SE and validation of equation, this model has robustness and practical applicability depending solely on descriptors derived from the chemical structures of N-(aryl)-2-thiophen-2-ylacetamide derivatives, which was further proved by cross validation tests. In conclusion, the developed QSPR models may be helpful in the development of new derivatives with improved lipophilicity.

REFERENCES

1. C. Dye, S. Scheele, P. Dolin, V. Pathania and M.C. Raviglione, *J. Am. Med. Assoc*. **282**, 677 (1999).
2. N.S. Shah, A. Wright, G.H. Bai, L. Barrera, F. Boulahbal, N.M. Casabona, F. Drobniewski, C. Gilpin, M. Havelkova, R. Lepe, R. Lumb, B. Metchock, F. Portaels, M.F. Rodrigues, S.R. Gerdes, A.V. Deun, V. Vincent, K. Laserson, C. Wells and J.P. Cegielski, *Emerg. Infect Dis*., **13**, 380 (2007).
3. H. Yu and A. Adedoyin, *Drug Discov. Today*, **8**, 852 (2003).
4. C.A. Lipinski, F. Lombardo, B.W. Dominy and P.J. Feeney, *Adv. Drug Deliv. Rev*., **46**, 3 (2001).
5. C. Hansch, J.P. Bjorkroth and A. Leo, *J. Pharm. Sci*., **76**, 663 (1987).
6. G.L. Amidon, S.H. Yalkowsky, S.T. Anik and S.C. Valvani, *J. Phys. Chem*., **79**, 2239 (1975).
7. M.C.S. Lourenco, F.R. Vicente, M.G.M.O. Henriques, A.L.P. Candea, R.S.B. Goncalves, T.C.M. Nogueira, M.L. Ferreira and M.V.N de Souza, *Bioorg. Med. Chem. Lett*., **17**, 6895 (2007).
8. C. Hansch and A. Leo, Substituents Constant for Correlation Analysis: In Analytical Chemistry and Biology, Wiley Interscience Publisher, p. 34 (1979).
9. http://www.statpages.org/miller/openstat/.

(*Received*: 3 October 2009; *Accepted*: 17 February 2010) AJC-8460