Academic Sciences Asian Journal of Pharmaceutical and Clinical Research
Vol 5, Issue 1, 2012 ISSN - 0974-2441

Research Article

ANTIMYCOBACTERIAL ACTIVITIES OF N-BENZYLSALICYLAMIDES AND N-BENZYLSALICYLTHIOAMIDES DERIVATIVES AGAINST *MYCOBACTERIUM TUBERCULOSIS* CNCTC MY (331/ 88) AND *MYCOBACTERIUM KANSASII* CNCTC MY (6509/ 96): QSAR MODELING USING ELECTROTOPOLOGICAL STATE ATOM (E-STATE) PARAMETERS

SUPRATIM RAY

Division of Pharmaceutical Chemistry, Dr. B C Roy College of Pharmacy & Allied Health Sciences, Bidhannagar, Durgapur, 713 206, India,
Email: supratimray_in@yahoo.co.in

Received: 09 October 2011, Revised and Accepted: 18 November 2011

ABSTRACT

This study gives a quantitative structure activity relationship (QSAR) correlation of the fourty four N-Benzylsalicylamides and N-Benzylsalicylthioamides derivatives properties reported by Dolezal et al against *Mycobacterium tuberculosis* CNCTC My (331/88) as well as *Mycobacterium kansasii* CNCTC My (6509/96). The reported minimum inhibitory concentrations [MIC] of the compounds determined after 21 days of incubation. The study was performed using electrotopological state atom (E-state) parameter as descriptors. Different statistical tools used in this communication are stepwise regression analysis and partial least squares analysis (PLS). For antitubercular activity considering internal (Q^2) and external validation (R^2_{pred}) (Q^2= 0.669, R^2_{pred}=0.774) PLS analysis was found to be the best. In case of activity against *Mycobacterium kansasii*, based on internal validation PLS analysis (Q^2= 0.6043) and on the basis of external validation stepwise regression analysis was found to be the best model (R^2_{pred}=0.7802).

Key words: QSAR, E-state, Stepwise regression, PLS, N-Benzylsalicylamides, Benzylsalicylthioamides, *Mycobacterium tuberculosis*, *Mycobacterium kansasii*

INTRODUCTION

TB is the world's second most cause of death due to an infectious disease, after acquired immune deficiency syndrome (AIDS) [1]. Drug-resistant strains of *Mycobacterium tuberculosis* (*M. tb*), in association with the deep synergy of TB with human immunodeficiency virus (HIV), are the major contributors for the disease. Excessive use of antibiotics is generally accepted to be the main reason for increased antibiotic resistance among bacteria [1, 2]. So the emergence of antibiotic-resistant pathogen agents is a serious health problem worldwide today. The time required for the treatment of TB is between 6 to 9 months. This long period leads to the lack of compliance, which in turn can be responsible for the relapse and emergence of MDR-TB strains [3]. The most common disease produced by of *Mycobacterium kansasii* infection is a chronic pulmonary infection that resembles pulmonary tuberculosis. However, it may also infect other organs. *M kansasii* infection is the second-most-common nontuberculous opportunistic mycobacterial infection associated with AIDS [4]. Due to emergence of multidrug resistance of the antitubercular drugs, there is an urgent need for the development of new drug candidate as well as gaining further (and deeper) knowledge of the mechanisms of action of existing (and future) active compounds. In this context a QSAR study was performed to the antitubercular activities of two moieties N-Benzylsalicylamides and N-Benzylsalicylthioamides derivatives reported by Dolezal et al [5] against *Mycobacterium tuberculosis* CNCTC My (331/88) as well as against *Mycobacterium kansasii* CNCTC My (6509/96).

MATERIALS AND METHODS

Electrotopological state atom (E-state) index

Structural specificity of a drug molecule is exhibited at an atomic or fragmental level instead of the whole molecule. In the drug receptor interaction phenomenon, a portion of the molecule (pharmacophore) may play more important role than the other segments. Though basic information for constitution of topological indices are derived from the atom level (count of atoms, bonds, paths of bonds, etc.), most of the indices are applied to the whole molecule after summing up all components over the whole molecule. Thus QSAR studies at the atomic or fragmental level are justified in the present context [6].

The electrotopological state atom (E-state) index developed by Hall and Kier [7] is an atom level descriptor encoding both the electronic character and topological environment of each skeletal atom in a molecule. The E-state of a skeletal atom is formulated as an intrinsic value I_i plus a perturbation term ΔI_i, arising from the electronic interaction within the molecular topological environment of each atom in the molecule.

The intrinsic value has been defined as the ratio of a measure of electronic state (Kier-Hall valence state electronegativity) to the local connectedness. The count of valence electrons which are the most reactive and involved in chemical reactions and bond formations are considered in the expression of *I* to encode the electronic feature. To reflect differences in electronegativity among the atoms, principal quantum number is employed in the expression of *I*. The topological attribute is included by using adjacency count of atom. The intrinsic value of an atom *i* is defined as

$$I_i = \left[(2/N)^2 \delta^v + 1 \right] / \delta \qquad \textbf{(i)}$$

In Eq. (i), N stands for principal quantum number and δ^v and δ indicate the count of valence electrons and sigma electrons associated with the atom *i* in the hydrogen suppressed graph. The intrinsic electrotopological state calculated according to Eq. (i) produces different values of an atom in different degrees of substitution (branching). The values are also different for different atoms having differences in electronegativity. The intrinsic values increase with increase in electronegativity or electron-richness and decrease with increase in branching (substitution).

The perturbation factor for the intrinsic state of atom *i* is defined as

$$\Delta I_i = \sum_{j \neq i} \frac{I_i - I_j}{r^2_{ij}} \qquad \textbf{(ii)}$$

In Eq. (ii) r_{ij} stands for the graph separation factor, i.e., count ofskeletal atoms in the shortest path connecting the atoms *i* and *j* including both atoms.

Summation of intrinsic state of an atom and influence of the field is called electrotopological state of the atom.

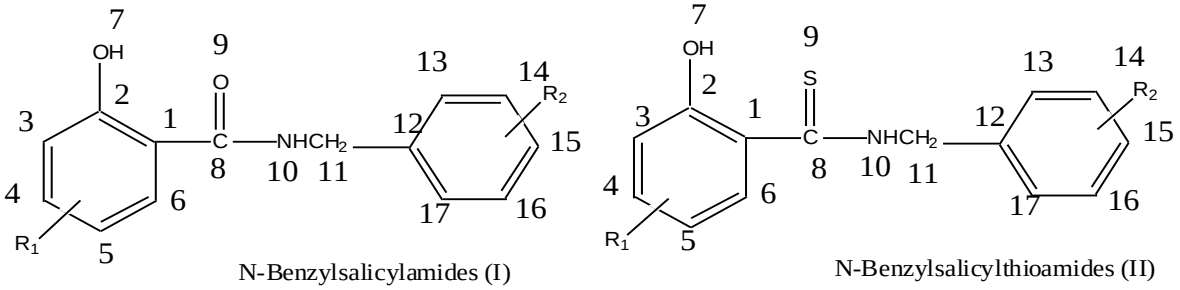
$$S_i = I_i + \sum_{j \neq i} \Delta I_{ij} \qquad \textbf{(iii)}$$

It is a representation of molecular structure information as it varies with changes in structural features including branching, cyclicity, homologation, heteroatom variation, and changes in relative positions of different groups. The electrotopological state considers both bonded and non-bonded interactions: the bonded component depends simply on differences in electronegativity among the adjacent atoms. The non-bonded interactions may be through inductive effect across the skeleton and is a function of graph separation factor and electronegativity differences. Thus, electrotopological state represents electronic distribution information modified by both local and global topology. The information encoded in the E-state value for an atom is the electronic accessibility at that atom.

Data treatment and software

The *in vitro* antimycobacterial activities of N-Benzylsalicylamides and N-Benzylsalicylthioamides derivatives against *Mycobacterium tuberculosis* CNCTC My (331/88) as well as against *Mycobacterium kansasii* CNCTC My (6509/96) reported by Dolezal et al [5] were used as the model data-set for the present QSAR analysis (**Table 1**). The reported minimum inhibitory concentrations [MIC] of the compounds determined after 21 days of incubation were in μM range which was converted to mM range and then to logarithmic scale [log (10^3 / MIC)]. The QSAR analysis was performed using electrotopological state atom (E-state) parameter. The whole data set contain fourty four compounds and all the compounds contain 17 common atoms (excluding hydrogen). The atoms of the molecules were numbered keeping serial numbers of the common atoms same in all the compounds (as shown in **Figure 1**). The electrotopological states of the 17 common atoms for all of the compounds were found out using a VISUAL BASIC program SRETSA developed partly by the author [8]. The program uses, as input, only the connection table in a specific format along with intrinsic state values of different atoms. To the output file thus obtained, the biological activity data were introduced to make it ready for subsequent regression analysis.

Figure 1: Common atom of the molecules



N-Benzylsalicylamides (I) N-Benzylsalicylthioamides (II)

Model development

To begin the model development process, the whole data set (n=44) was divided into training (n=33, 75% of the total number of compounds) and test (n=11, 25% of the total number of compounds) sets by *k*-means clustering technique [9] applied on standardized descriptor matrix of the E-state parameters. QSAR models were developed using the training set compounds (optimized by Q^2), and then the developed models were validated (externally) using the test set compounds. The stepwise regression and PLS analysis were performed using statistical software MINITAB [10].

Stepwise Regression

In stepwise regression [11], a multiple term linear equation was built step-by-step. The basic procedures involve (1) identifying an initial model, (2) iteratively "stepping", i.e., repeatedly altering the model of the previous step by adding or removing a predictor variable in accordance with the "stepping criteria", (F = 4 for inclusion; F = 3.9 for exclusion) in our case and (3) terminating the search when stepping is no longer possible given the stepping criteria, or when a specified maximum number steps has been reached. Specifically, at each step all variables are reviewed and evaluated to determine which one will contribute most to the equation. That variable will then be included in the model, and the process started again. A limitation of the stepwise regression search approach is that it presumes that there is a single "best" subset of X variables and seeks to identify it. There is often no unique "best" subset, and all possible regression models with a similar number of X variables as in the stepwise regression solution should be fitted subsequently to study whether some other subsets of X variables might be better.

PLS

PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables [12, 13]. It gives a reduced solution, which is statistically more robust than MLR. The linear PLS model finds "new variables" (latent variables or X scores) which are linear combinations of the original variables. To avoid over fitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are nonsignificant. Application of PLS thus allows the construction of larger QSAR equations while still avoiding over fitting and eliminating most variables. PLS is normally used in combination with cross validation to obtain the optimum number of components. This ensures that the QSAR equations are selected based on their ability to predict the data rather than to fit the data. In case of PLS analysis on the present data set, based on the standardized regression coefficients, the variables with smaller coefficients were removed from the PLS regression until there was no further improvement in Q^2 value irrespective of the components.

Statistical qualities

The statistical qualities of the equations were judged by the parameters such as *determination coefficient* (R^2) and *variance ratio* (F) at specified *degrees of freedom* (df) [14]. The generated QSAR equations were validated by leave-one-out *cross-validation* R^2 (Q^2) and *predicted residual sum of squares* (*PRESS*) [14-16] and then were used for the prediction of antimycobacterial activity of the test set compounds. The prediction qualities of the models were judged by statistical parameters like predictive R^2 (R^2_{pred}).

RESULTS AND DISCUSSION

Membership of compounds in different clusters generated using *k*-means clustering technique is shown in **Table 2**. The test set size was set to approximately 25% to the total data set size [17] and the test set members along with their observed and calculated activity are given in **Table 3**. Statistical qualities of all important models are listed in **Table 4**. The results obtained from different statistical methods are described below and the interpretations of the equations are also depicted.

Activity against *Mycobacterium tuberculosis*

Stepwise regression

Using stepping criteria based on F value (F = 4.0 for inclusion; F = 3.9 for exclusion), the following equation was obtained with E-State parameters.

$$pC_{21d} = 1.786(\pm 0.089) + 1.73(\pm 0.196)S_8$$
$$n_{Training} = 33, R^2 = 0.716, R_a^2 = 0.707, S = 0.450, F = 78.16(df\,1,31), \quad (1)$$
$$Q^2 = 0.681, PRESS = 7.04, n_{Test} = 11, R^2_{pred} = 0.535$$

The standard errors of the respective E-state indices are mentioned within parentheses. Eq. (1) could explain 70.7% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 68.10%. While Eq. (1) was applied for prediction of test set compounds, the predictive R^2 value for the test set was found to be 0.535. The positive coefficient of S_8 indicates that activity increases with increase in E-state value of atom 8. Position 8 indicates the importance of carbon atom bearing the hetero atom oxygen or sulphur. Compounds like **17, 36, 37 and 38** with high values of E-state parameter for atom 8 showed comparative higher activity.

PLS

The number of optimum components was 2 to obtain the final equation (optimized by cross validation). Based on the standardized regression coefficients, the following variables were selected for the final equation:

$$pC_{21d} = 0.7158 + 0.7323S_1 + 0.8828S_2 + 0.3833S_8$$
$$-0.05211S_9 + 0.6194S_{10} + 0.2632S_{11} - 0.2035S_{13} - 0.1682S_{14}$$
$$n_{Training} = 33, R^2 = 0.739, R_a^2 = 0.721, Q^2 = 0.669, S = 0.037$$
$$PRESS = 7.32, F = 42.57(df\ 2,30), n_{Test} = 11, R^2_{pred} = 0.774$$

(2)

Eq. (2) could explain 72.1% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 66.90%. While Eq. (2) was applied for prediction of test set compounds, the predictive R^2 value for the test set was found to be 0.774. The negative coefficients of S_9, S_{13} and S_{14} indicate that activity decreases with increase in E-state value of atoms 9, 13 and 14 respectively. Compounds with high values of E-state parameter for atom 9 (S_9) (like **10 and 15)** for atom 13 (S_{13}) (like **1, 10 and 33)** and for atom 14 (S_{14}) (like **1, 7, 8, 10 and 13)** showed comparatively poor activity. The hetero atoms either oxygen or sulphur at position 9 is negatively contributed towards activity.

The positive coefficient of S_1, S_2, S_8, S_{10}, and S_{11} indicates that activity increases with increase in E-state value of atom 1, 2, 8, 10 and 11 respectively. Compounds with high values of E-state parameter for atom 1 (S_1) (like **17, 36, 37 and 38**) for atom 2 (S_2) (like **22, 36** and **38)** for atom 10 (S_{10}) (like **17, 20, 22, 36 and 37)** and for atom 11 (S_{11}) (like **16, 17, 36, 37 and 38)** showed comparatively higher activity. Position 1 indicates the importance of the connecting moiety methylcarboxamido / methylthiocarboxamido group between two substituted phenyl groups. The carbon atom bearing the hydroxyl group is positively contributed towards activity. Position 10 and 11 indicates the positive contribution of amino and methyl group of the connecting moiety between two substituted phenyl groups towards activity.

Activity against *Mycobacterium kansasii*

Stepwise regression

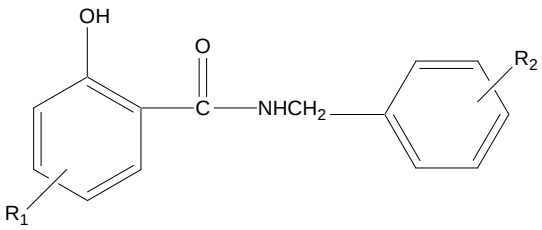
Using stepping criteria based on F value (F = 4.0 for inclusion; F = 3.9 for exclusion), the following equation was obtained with E-State parameters.

$$pC_{21d} = 0.425(\pm 0.210) + 3.13(\pm 0.410)S_1$$
$$n_{Training} = 33, R^2 = 0.652, R_a^2 = 0.641, S = 0.452, F = 58.11(df\ 1,31),$$
$$Q^2 = 0.6014, PRESS = 7.27, n_{Test} = 11, R^2_{pred} = 0.7802$$

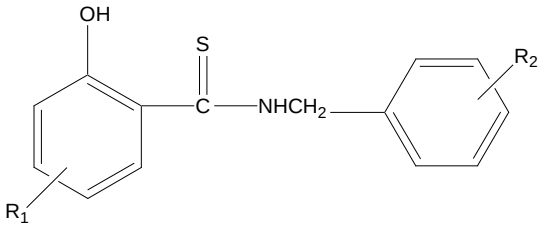
(3)

The standard errors of the respective E-state indices are mentioned within parentheses. Eq. (3) could explain 64.10% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 60.14%. While Eq. (3) was applied for prediction of test set compounds, the predictive R^2 value for the test set was found to be 0.7802. Position 1 indicates the importance of the connecting moiety methylcarboxamido / methylthiocarboxamido group between two substituted phenyl groups. Compounds with high values of E-state parameter for atom 1 (like **18, 22, 24, 36, 37 and 38)** showed comparatively higher activity.

Table 1: Molecular scaffolds of the compounds along with their activity.



N-Benzylsalicylamides (I)



N-Benzylsalicylthioamides (II)

Compound No.	Type of compound	R_1	R_2	MIC value (µmol/L) against *Mycobacterium tuberculosis* for 21 days (C_{21d})	pC_{21d} = Log ($1000/C_{14d}$)	MIC value (µmol/L) against *Mycobacterium kansasii* for 21 days (C_{21d})	pC_{21d} = Log ($1000/C_{14d}$)
1	I	H	4-tert-but	32	1.49485	62.5	1.20412
2	I	H	3-CF_3	62.5	1.20412	62.5	1.20412
3	I	5-Br	3-Br	32	1.49485	32	1.49485
4	I	5-Br	4-Br	32	1.49485	32	1.49485
5	I	3,5 Cl_2	4-tert-	32	1.49485	62.5	1.20412

			but				
6	I	4-Cl	4-Br	32	1.49485	32	1.49485
7	I	4-CH_3	H	125	0.90309	125	0.90309
8	I	4-CH_3	4-CH_3	500	0.30103	500	0.30103
9	I	4-CH_3	4-Cl	125	0.90309	250	0.60206
10	I	4-CH_3	4-tert-but	62.5	1.20412	62.5	1.20412
11	I	4-CH_3	3-NO_2	62.5	1.20412	125	0.90309
12	I	4-OCH_3	3-Cl	125	0.90309	125	0.90309
13	I	3-CH_3	H	125	0.90309	125	0.90309
14	I	3-CH_3	4-Cl	32	1.49485	62.5	1.20412
15	I	3,5 Br_2	4-CF_3	62.5	1.20412	62.5	1.20412
16	II	H	H	1	3	2	2.69897
17	II	H	4-CH_3	0.5	3.30103	1	3
18	II	H	4-Cl	2	2.69897	1	3
19	II	H	4-OCH_3	4	2.39794	4	2.39794
20	II	H	3,4 Cl_2	0.98	3.008774	4	2.39794
21	II	H	4-F	1	3	2	2.69897
22	II	H	3-CH_3	1	3	2	2.69897
23	II	H	4-tert-but	2	2.69897	2	2.69897
24	II	H	3-Cl	1	3	1	3
25	II	H	3-CF_3	1	3	2	2.69897
26	II	5-Br	3,4 Cl_2	4	2.39794	16	1.79588
27	II	5-Br	3-Br	2	2.69897	8	2.09691
28	II	5-Br	4-Br	2	2.69897	4	2.39794
29	II	5-Cl	H	2	2.69897	8	2.09691
30	II	5-Cl	3,4 Cl_2	4	2.39794	16	1.79588
31	II	5-Cl	4-F	2	2.69897	8	2.09691
32	II	3,5 Cl_2	3,4 Cl_2	32	1.49485	32	1.49485
33	II	3,5 Cl_2	4-tert-but	32	1.49485	62.5	1.20412
34	II	4-Cl	4-Br	2	2.69897	8	2.09691
35	II	4-CH_3	H	1	3	2	2.69897
36	II	4-CH_3	4-CH_3	0.25	3.60206	1	3
37	II	4-CH_3	4-Cl	0.5	3.30103	1	3
38	II	4-CH_3	4-tert-but	0.5	3.30103	1	3
39	II	4-CH_3	3-NO_2	2	2.69897	4	2.39794
40	II	5-OCH_3	H	8	2.09691	16	1.79588

Table 2: *k*-Means clustering of compounds using standardized descriptors.

Cluster No.	No. of compounds in different clusters	Compounds (Sl nos.) in each clusters
1	19	16 17 18 19 22 23 24 27 28 29 34 35 36 37 38 40 41 42 43
2	5	2 11 15 25 44
3	12	1 3 4 5 6 7 8 9 10 12 13 14
4	8	20 21 26 30 31 32 33 39

Table 3: Observed and calculated antitubercular activities from different models

Sl. No.	Observed activity [a] against *Mycobacterium tuberculosis* (pC_{21d})	Cal[b]	Cal[c]	Observed activity [d] against *Mycobacterium kansasii* (pC_{21d})	Cal[e]	Cal[f]
			Training set			
1	1.49485	1.330584	1.312883	1.20412	1.338099	1.338099
2	1.20412	0.94614	1.038939	1.20412	1.134878	1.134878
4	1.49485	1.272128	1.258764	1.49485	1.27033	1.27033
5	1.49485	1.064113	0.791689	1.20412	0.57751	0.57751
6	1.49485	1.216714	1.11478	1.49485	1.12244	1.12244
8	0.30103	1.338433	1.35027	0.30103	1.268042	1.268042
9	0.90309	1.281369	1.325477	0.60206	1.246337	1.246337
10	1.20412	1.337562	1.338562	1.20412	1.167846	1.167846
13	0.90309	1.34157	1.397429	0.90309	1.17393	1.17393
14	1.49485	1.287238	1.361699	1.20412	1.109535	1.109535
15	1.20412	0.868031	0.911726	1.20412	0.6388	0.6388
16	3	2.770247	2.706584	2.69897	2.573256	2.573256
18	2.69897	2.715915	2.670723	3	2.508849	2.508849
19	2.39794	2.725383	2.672798	2.39794	2.480629	2.480629
20	3.008774	2.64559	2.821296	2.39794	2.580733	2.580733
22	3	2.774455	2.832647	2.69897	2.638171	2.638171
24	3	2.699923	2.848711	3	2.638427	2.638427
25	3	2.28761	2.615203	2.69897	2.34517	2.34517
26	2.39794	2.606536	2.769658	1.79588	2.424993	2.424993
27	2.69897	2.717858	2.766178	2.09691	2.476773	2.476773
28	2.69897	2.71365	2.634514	2.39794	2.366661	2.366661
30	2.39794	2.512354	2.615389	1.79588	2.212686	2.212686
31	2.69897	2.477554	2.418584	2.09691	2.100377	2.100377
33	1.49485	2.505635	2.177475	1.20412	1.682314	1.682314
34	2.69897	2.658237	2.48925	2.09691	2.217692	2.217692
35	3	2.777224	2.733534	2.69897	2.404173	2.404173
36	3.60206	2.779956	2.724742	3	2.363293	2.363293
38	3.30103	2.779084	2.718245	3	2.267498	2.267498
39	2.69897	2.501448	2.821479	2.39794	2.36142	2.36142
40	2.09691	2.663434	2.54832	1.79588	2.189259	2.189259
41	2.39794	2.693491	2.553371	2.09691	2.164156	2.164156
43	2.39794	2.72876	2.737451	2.39794	2.205868	2.205868
44	1.49485	2.309553	2.297513	1.20412	1.743605	1.743605
			Test set			
3	1.49485	1.267398	1.409223	1.49485	1.240522	1.412206
7	0.90309	1.335702	1.361206	0.90309	1.392011	1.310732
11	1.20412	1.051757	1.365333	0.90309	0.985027	1.197211
12	0.90309	1.200769	1.372824	0.90309	1.134603	1.26351
17	3.30103	2.772977	2.701914	3	2.531303	2.530551
21	3	2.61079	2.625047	2.69897	2.299451	2.468865
23	2.69897	4.246148	2.870603	2.69897	2.271072	2.613198
29	2.69897	2.637011	2.497846	2.09691	2.165162	2.202944
32	1.49485	2.379118	2.310862	1.49485	1.625919	1.829172
37	3.30103	2.722892	2.699948	3	2.472969	2.341588
42	3	2.623166	2.698618	3	2.236363	2.231821

[a] Observed activity (ref. 5); [b] Calculated from eq. (1); [c] Calculated from eq. (2); [d] Observed activity (ref. 5); [e] Calculated from eq. (3); [f] Calculated from eq. (4);

Table 4: Statistical comparison of different models

Type of statistical methods	*Mycobacterium tuberculosis*				*Mycobacterium kansasii*			
	R^2	R_a^2	Q^2	R^2_{pred}	R^2	R_a^2	Q^2	R^2_{pred}
Stepwise	0.716	0.707	0.681	0.535	0.652	0.641	0.6014	**0.7802**
PLS	**0.739**	**0.721**	**0.669**	**0.774**	**0.7047**	**0.685**	**0.6043**	0.7793

***The best values of different parameters are shown in bold**

PLS

The number of optimum components was 2 to obtain the final equation (optimized by cross validation). Based on the standardized regression coefficients, the following variables were selected for the final equation:

$$pC_{21d} = 10.6365 + 0.7505S_1 + 1.0577S_6 - 1.2217S_7$$
$$+0.3424S_8 - 0.044S_9 + 0.5170S_{10} - 0.1334S_{14}$$
$$n_{Training} = 33, R^2 = 0.7047, R_a^{\ 2} = 0.685, Q^2 = 0.6043, S = 0.032$$
$$PRESS = 7.22, F = 35.80(df\ 2,30), n_{Test} = 11, R^2_{pred} = 0.7793$$

(4)

Eq. (4) could explain 68.5% of the variance (adjusted coefficient of variation) and leave – one – out predicted variance was found to be 60.43%. While Eq. (4) was applied for prediction of test set compounds, the predictive R^2 value for the test set was found to be 0.7793. The negative coefficients of S_7, S_9 and S_{14} indicate that activity decreases with increase in E-state value of atoms 7, 9 and 14 respectively. Compounds with high values of E-state parameter for atom 7 (S_7) (like **5**, **10, 14, 15, 33 and 44)** for atom 9 (S_9) (like **1, 7, 8, 9, 11 and 12)** and for atom 14 (S_{14}) (like **1, 7, 8, 10, 13 and 33)** showed comparatively poor activity. The hetero atoms either oxygen or sulphur at position 9 is negatively contributed towards activity. The hydroxyl group at position 7 is also negatively contributed towards activity. The positive coefficient of S_1, S_6, S_8 and S_{10} indicates that activity increases with increase in E-state value of atom 1, 6, 8 and 10 respectively. Compounds with high values of E-state parameter for atom 6 (S_6) (like **17, 36, 37 and 38**) for atom 8 (S_8) (like **17, 18, 23, 36** and **38)** and for atom 10 (S_{10}) (like **17, 22, 36 and 38)** showed comparatively higher activity. Position 8 indicates the importance of carbon atom bearing the hetero atom oxygen or sulphur. Position 10 indicates the positive contribution of amino group of the connecting moiety between two substituted phenyl groups towards activity.

OVERVIEW AND CONCLUSIONS

The whole dataset (n=44) was divided into a training set (33 compounds) and a test set (11 compounds) based on *k*-means clustering of the standardized descriptor matrix and models were developed from the training set. The predictive ability of the models was judged from the prediction of the activity of the test set compounds. For antitubercular activity considering internal (Q^2) and external validation (R^2_{pred}) (Q^2= 0.669, R^2_{pred}=0.774) PLS analysis was found to be the best. In case of activity against *Mycobacterium kansasii*, based on internal validation PLS analysis (Q^2= 0.6043) and on the basis of external validation stepwise regression analysis was found to be the best model (R^2_{pred}=0.7802). For both the mycobacterium species the developed models indicate the importance of connecting moiety methylcarboxamido / methylthiocarboxamido group between two substituted phenyl groups. It was also found that there is positive contribution of amino group of the connecting moiety between two substituted phenyl groups towards activity. The carbon atom at position 8 bearing the hetero atom oxygen or sulphur is also positively contributed towards activity. But the hetero atoms either oxygen or sulphur at position 9 is negatively contributed towards activity.

REFERENCES

1. Guerrin-Tran E, Thiolet JM, Rousseau C, Henry S, Poirier C, Che D, Vinas JM, Jarlier V, Robert J. An evaluation of data quality in a network for surveillance of Mycobacterium tuberculosis resistance to antituberculosis drugs in Ile-de-France region-2001-2002. Eur J Epidemiol 2006**;** 21: 783-785.
2. Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C. Tuberculosis. Lancet 2003; 362: 887-899.
3. Matsumoto M, Hashizume H, Tsubouchi H, Sasaki H, Itotani M, Kuroda H, Tomishige T, Kawasaki M, Komatsu M. Screening for novel antituberculosis agents that are effective against multidrug resistant tuberculosis**.** Curr Top Med Chem 2007; 7: 499-507.
4. Bloch KC, Zwerling L, Pletcher MJ, Hahn JA, Gerberding JL, Ostroff SM. Incidence and clinical implications of isolation of Mycobacterium kansassi: results of a 5-year, population-based study. Ann Int Med 1998; 129: 698– 704.
5. Dolezal R, Waisser K, Petrlikova E, Kunes J, Kubicova L, Machacek M, Kaustova J, Martin Dahse H. N Benzylsalicylthioamides: Highly active potential Antituberculotics. Arch Pharm Chem Life Sci 2009; 342: 113-119.
6. Hall LH, Mohney B, Kier LB. The Electrotopological State: An Atom Index for QSAR. Quant Struct- Act Relat 1991; 10: 43-51.
7. Kier LB, Hall LH. An Electrotopological State Index for Atoms in Molecules. Pharm Res 1990; 7: 801-807.
8. SRETSA is statistical software in Visual Basic, developed by Ray S, Biswas R. and standardized using known data sets.
9. Leonard JT, Roy K. On Selection of Training and Test Sets for the Development of Predictive QSAR models. QSAR Comb Sci 2006; 25: 235-251.
10. MINITAB is statistical software of Minitab Inc, USA, http://www.minitab.com.
11. Darlington RB. Regression and linear models. New York: McGraw Hill; 1990.
12. Wold S. PLS for multivariate linear modeling. In: Van de Waterbeemd H, editor. Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry). Weinheim: VCH; 1995. p. 195-218.
13. Fan Y, Shi LM, Kohn KW, Pommier Y, Weinstein JN. Quantitative structure-antitumor activity relationships of camptothecin analogs: Cluster analysis and genetic algorithm-based studies. J Med Chem 2001; 44: 3254-3263.
14. Snedecor GW, Cochran WG. Statistical Methods. New Delhi: Oxford & IBH Publishing Co. Pvt. Ltd.; 1967.
15. Debnath AK. In: Ghose AK, Viswanadhan VN, editors. Combinatorial library design and evaluation: Principles, software tools, and applications in drug discovery. New York: Marcel Dekker; 2001. p. 73–129.
16. Roy K. On some aspects of validation of predictive QSAR models. Expert Opin Drug Discov 2007; 2: 1567-1577.
17. Roy PP, Leonard JK, Roy K. Exploring the impact of the size of training sets for the development of predictive QSAR models. Chemom Intell Lab Sys 2008; 90: 31-42.