Journal of Enzyme Inhibition and Medicinal Chemistry, June 2009; 24(3): 850–858

Insights through AM1 calculations into the structural requirement of *N*-hydroxythiosemicarbazone analogs as anti-tubercular agents

REVATHI A. GUPTA, ARUN KUMAR GUPTA, LOVE KUMAR SONI, & S. G. KASKHEDIKAR

Molecular Modelling Study Group, CADD Laboratory, Computational Chemistry Research, Department of Pharmacy, Shri G. S. Institute of Technology and Science, 23 Park Road, Indore 452003, MP, India

(*Received 26 April 2008; accepted 8 August 2008*)

Abstract

Fujita-Ban, Hansch substituent constants, topological descriptors and conformational dependent descriptors were explored for quantification of anti-tubercular activity of *N*-hydroxythiosemicarbazone analogs. All the approaches gave statistically sound model which accounts for more than 88% of the explain variance against anti-tubercular activity except Fujita-Ban (~75%). Fujita-Ban & Hansch approaches having certain limitation, however, another approache showed their significant role in explaining activity of the modified scaffolds. QSAR study of *N*-Hydroxythiosemicarbazone analogs furnished some important structural insights i.e., the R position is more prone for improving inhibitory activity and the R_1 position play a decisive role in the ionic interaction of the ligand with macromolecules. On the basis of findings, *N*-hydroxythiosemicarbazones interaction model with macromolecule of *M. Tuberculosis* has been proposed. These interactions might be helpful in further development of potent anti-tubercular agents.

Keywords: *Fujita-Ban approach, Hansch substituent constant approach, QSAR, anti-tubercular,* N-*Hydroxythiosemicarbazone analogs*

Introduction

Tuberculosis (TB) is responsible for more morbidity in humans than any other single infectious organism in the world. Almost one-third of the world's population has been infected with the causative organism *Mycobacterium tuberculosis*, eight million become sick with TB every year, and annually it accounts for approximately three million deaths [1]. The increase of TB during recent years was largely due to poverty, overcrowding, globalization and the synergy between HIV and TB [2–4]. Moreover, Long-term therapies frequently led to patient non-compliance and, in turn, contributed to the emergence of multi-drug resistant TB (MDR-TB) [5].

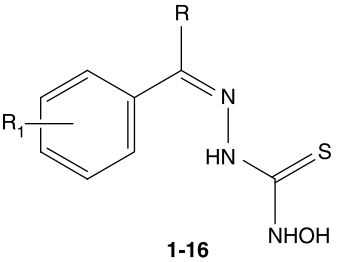
Only few new antibiotics against TB have been developed in the past 40 years [6]. Despite of this, an urgent need for new antibiotics towards TB is widely acknowledged and a Global Alliance for TB Drug Development has been established with goals that include increasing compliance with current therapies, shortening treatment times with the development of new adjuvant drugs, improved treatment for multi-drug resistant TB and new treatments for persistent *M. tuberculosis*.

The QSAR analysis of the anti-tubercular agents is the recent interest area of the researchers. The QSAR was performed on a series of *N*-Hydroxythiosemicarbazone analogs [7]. The emphasis was focused on the quantification of structure activity relationship with a view to delineate the influence of key physicochemical properties on mycobacterial inhibitory activity, which will aid in the designing of potent and safer inhibitors. The quantification of physicochemical properties was done with the help of regression technique.

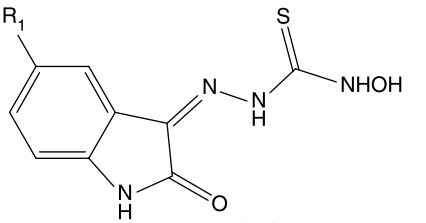
Correspondence: Dr. S.G. Kaskhedikar, Molecular Modelling Study Group, CADD Laboratory, Computational Chemistry Research, Department of Pharmacy, Shri G.S. Institute of Technology and Science, 23 Park Road, Indore 452003, M.P., India.

ISSN 1475-6366 print/ISSN 1475-6374 online © 2009 Informa UK Ltd.
DOI: 10.1080/14756360802421094

Table I. Structure and activity of *N*-hydroxythiosemicarbazone analogs.



1-16



17-18

Comd. No.	R	R_1	MIC⋆	pMIC†
1	H	H	64.02	4.194
2	H	2-OH	29.58	4.529
3	H	4-OH	29.58	4.529
4	H	4-$N(CH_3)_2$	52.45	4.280
5	H	4-OCH_3	14.95	4.825
6	H	4-CH_3	12.97	4.887
7	H	3-OCH_3,4-OH	27.21	4.565
8	H	4-Cl	13.02	4.885
9	H	4-NO_2	6.49	5.188
10	H	2-NO_2	6.49	5.188
11	CH_3	H	14.95	4.825
12	CH_3	2-OH	13.89	4.857
13	CH_3	4-CH_3	6.98	5.156
14	CH_3	4-Cl	3.2	5.495
15	CH_3	4-NO_2	0.78	6.108
16	C_6H_5	4-Br	0.28	6.553
17	–	H	64.02	4.194
18	–	F	24.58	4.609

⋆Minimum inhibitory concentration in μM against *Mycobacterium tuberculosis* H37Rv.; †Negative logarithm of minimum inhibitory concentration in mole

Experimental

The *M. tuberculosis* minimum inhibitory concentration data of *N*-Hydroxythiosemicarbazone analogs was taken from the reported work of Sriram et al. [7] (Table I). The inhibitory activity data (MIC in μM) was converted to negative logarithmic dose in mole (pMIC) because a QSAR is a linear free energy relationship, and from the Van't Hoff isotherm, free energy change during a process is proportional to the logarithm of the rate or equilibrium constant of the process ($\Delta G = -2.303\ RT \log K$).

Initially the series was subjected to Fujita-Ban analysis using regression technique in order to estimate the *de novo* contribution of substituents to the activity of the scaffold. Further, the Hansch approach was carried out to establish correlation between tuberculosis inhibitory activity and various substituent constants at position R and R_1 of the molecule (Table I). Values of substituent constants like hydrophobic (π), hydrophobic fragmental constants (F_{hyd}), steric (Molar refractivity or *MR*), hydrogen acceptor (*HA*), hydrogen donor (*HD*), and electronic descriptor like field effect ($\mathscr{F}$), resonance effect ($\mathscr{R}$) and Hammet's constant (σ) were taken from the reported data [8,9] and Verloop parameters (value of shape of each substituent) like *L*, B_1, B_2, B_3, B_4 were taken from reported work of Skagerberg et al. [10].

The molecular modeling study was performed using CS ChemOffice [11] and Dragon [12] program while the regression analysis was carried out on VALSTAT [13]. Structure of all the compounds was sketched out by builder module of the program. The sketched structures were subjected to energy minimization using molecular mechanics (MM2) until the root mean square (RMS) gradient value became smaller than 0.1 kcal/mol Å. The energy-minimized molecules were subjected to re-optimization via Austin model-1 (AM1) method until the RMS gradient attained a value smaller than 0.0001 kcal/mol Å using MOPAC. The geometry optimization of the lowest energy structure was carried out using EF routine. The minimized molecule was saved as MOL file format. The MOL file was used for calculation of various physicochemical properties with the help of Dragon program.

The data was transferred to the statistical program in order to establish a correlation between physicochemical parameters as independent variable and pMIC as dependent variable employing sequential multiple linear regression analysis method. In sequential multiple linear regression, the program searches

for all the permutation and combination sequentially for the given data set. The ± data within the parentheses is the standard deviation, associated with the coefficient of descriptor in regression equation. The various statistically significant equations were taken in consideration on the basis of observed squared correlation coefficient (r^2), the standard error of estimate (*SEE*) and the sequential Fischer test (*F*). The internal predictive powers of the equations were validated by leave-one-out (*loo*) [14,15] and leave-n-out (*lno*) cross-validation method considering predicted residual sum of square (*PRESS*), total sum of squares (*SSY*), cross-validated squared correlation coefficient (q^2 and q^2_{nout}), and standard deviation error of prediction (S_{DEP}). Bootstrapping analysis was performed to ascertained robustness and statistical confidence against the model. Bootstrapping squared correlation coefficient (r^2_{bs}) is average squared correlation coefficient of subset of compounds used in regression. Chances of fortuitous correlation were tested with the help of *Chance* statistics. Outliers investigated with the help of Z-score value.

Results and discussion

Fujita-Ban analysis was carried out in order to ascertain *de novo* contribution of the substituent group on the activity (Table II). In the Fujita-Ban expression only statistical significant group's contribution was considered. The significant trivalent expression was depicted in Equation (1) (model-1).

$$\begin{aligned}\text{pMIC} = {} & 0.505(\pm 0.175)\text{R} - \text{CH}_3 \\ & + 1.922(\pm 0.334)\text{R} - \text{C}_6\text{H}_5 \\ & + 0.765(\pm 0.243)\text{R}_1 - 4\text{NO}_2 + 4.631 \\ n = 16,\ & r = 0.892,\ r^2 = 0.795,\ SEE = 0.317, \\ F = 15.508,\ & F_{\max} = 11.145\end{aligned} \tag{1}$$

Fujita-Ban analysis of tuberculosis inhibitory activity of *N*-Hydroxythiosemicarbazone analogs inferred that the substitutions at 4th position of phenyl ring are favorable as compare to 2nd and 3rd positions (Table II). Especially substitution of nitro moiety at 4th position of phenyl ring result in potent inhibition as compared to other substituted analogs. *De novo* techniques also suggested that the R position is more prone for improving inhibitory activity. In the series substitution of methyl or phenyl moiety on this position are results in improvement in the activity.

Fujita-Ban expression gave insight to some important structural features i.e., nitro group on R_1 position is optimal for the activity and might be responsible for electronic interaction. Presence of the phenyl moiety on carbimino terminal (R position) might be helpful in accommodating the ligand in hydrophobic pocket more effectively as compared to un-substituted or methyl substituted analogs. The R position offer wide range of chemical space for further enhancing the inhibitory activity of core structure.

In extension of our study towards structural requirements, Hansch approach was explored through sequential multiple linear regression (SEQ-MLR) technique. The multivariant expressions were developed on the basis of adjustable correlation coefficient (r^2_{adj}). This parameter expresses the statistical significance of incorporated physicochemical descriptor in regression expression. Adjustable correlation coefficient is a measure of the % explained variation in the dependent variable that takes into account the relationship between the number of cases and the number of independent variable in the regression model. Although r^2 value should be increase if an independent variable is added to the expression, while r^2_{adj} value would be decrease if the added variable does reduce the unexplained variation enough to offset the loss of degrees of freedom. The SEQ-MLR furnished uni and bi-variant expression with increasing correlation coefficient (Equations (1) and (2)) and also the r^2_{adj} value is increasing significantly from uni to bi-variant expression.

$$\begin{aligned}\text{pMIC} = {} & 1.098(\pm 0.227)F_{hyd} + 4.451 \\ n = 16,\ & r = 0.791,\ r^2_{adj} = 0.599,\ SEE = 0.397, \\ F = 23.421,\ & F_{\max} = 11.946\end{aligned} \tag{2}$$

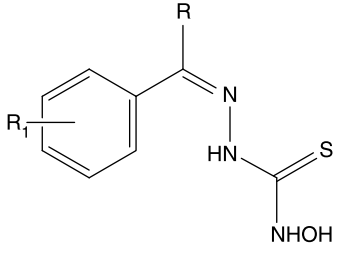
$$\begin{aligned}\text{pMIC} = {} & 0.071(\pm 0.012)MR + 0.734(\pm 0.167) \\ & \times \sigma_1 + 4.643 \quad n = 16,\ r = 0.915, \\ r^2_{adj} = 0.813,\ & SEE = 0.271,\ F = 33.655, \\ F_{\max} = 10.066 & \end{aligned} \tag{3}$$

Therefore we plan to move for higher variant expression like tri-variant. In case of tri-variant expression adjustable r^2 value significantly improved, hence these expressions (explained variance ~ 90%) were further explored for QSAR study. The proposed models should have statistical quality as well as predictive power therefore the expression was tested for their corroboration.

$$\begin{aligned}\text{pMIC} = {} & 0.942(\pm 0.121)F_{hyd} + 0.048(\pm 0.015)MR_1 \\ & + 0.795(\pm 0.132)\sigma_1 + 4.148 \quad n = 16, \\ r = 0.955,\ & r^2_{adj} = 0.891,\ SEE = 0.207, \\ F = 41.803,\ & F_{\max} = 9.831\end{aligned} \tag{4}$$

Equation (4) (model-2) shows better correlation coefficient ($r = 0.955$), which account for ~90% of the explained variance in the activity calculated as $r^2_{adj} = r^2(1 - 1/F)$ that depict in percentage when multiplied by 100. The data showed overall internal statistical significance level better than 99.9% as the

Table II. Fujita-Ban matrix of *N*-hydroxythiosemicarbazone analogs with calculated (Cal) value.



		R						R_1						
Comp. No.	μ*	CH_3	C_6H_5	2-OH	2-NO_2	3-OCH_3	4-OH	4-$N(CH_3)_2$	4-OCH_3	4-CH_3	4-Cl	4-NO_2	4-Br	Calc pMIC†
1	1	0	0	0	0	0	0	0	0	0	0	0	0	4.631
2	1	0	0	1	0	0	0	0	0	0	0	0	0	4.631
3	1	0	0	0	0	0	1	0	0	0	0	0	0	4.631
4	1	0	0	0	0	0	0	1	0	0	0	0	0	4.631
5	1	0	0	0	0	0	0	0	1	0	0	0	0	4.631
6	1	0	0	0	0	0	0	0	0	1	0	0	0	4.631
7	1	0	0	0	0	1	1	0	0	0	0	0	0	4.631
8	1	0	0	0	0	0	0	0	0	0	1	0	0	4.631
9	1	0	0	0	0	0	0	0	0	0	0	1	0	5.395
10	1	0	0	0	1	0	0	0	0	0	0	0	0	4.631
11	1	1	0	0	0	0	0	0	0	0	0	0	0	5.135
12	1	1	0	1	0	0	0	0	0	0	0	0	0	5.135
13	1	1	0	0	0	0	0	0	0	1	0	0	0	5.135
14	1	1	0	0	0	0	0	0	0	0	1	0	0	5.135
15	1	1	0	0	0	0	0	0	0	0	0	1	0	5.900
16	1	0	1	0	0	0	0	0	0	0	0	0	1	6.553

* μ is the value for core nucleus used in Fujita-Ban analysis; † Calculated pMIC considering Equation (1).

calculated variance ratio i.e., Fischer value (F) exceeded the tabulated $F_{(3,12\alpha 0.001)} = 12.7$. Fischer value suggested that the equations are applicable for more than 999 out of 1000 times. Although tabulated *F-value* are able to explained significance of MLR model, which obtained from same set of variables. If, However, MLR expression developed from subset of a large set of variables, then that might be affected by selection bias. In such case F_{max} value helpful in explanatory of unbiased expression. F_{max} values are dependent on the size of the descriptor pool. The standard F_{max} value (95% confidence interval) was calculated from power function reported by Livingstone et al. [16]. $F_{max} = (29.96n^{3.18}N^{0.207}/p^{0.82})\, e^{\ln(V2)[1.057\ln(V2) - 0.97\ln(n) - 3.97]}$ where N is the number of possible regression models, p is the size of the models, n is the number of cases generated for the random variables, and V2 is the second degree of freedom of the F-statistics i.e. $n - p - 1$. In the above mentioned model, F value exceeded the $F_{max(3,12\alpha 0.05)} = 9.831$. The orthogonality of the substituent in SEQ-MLR was established through variance inflation factor values (*VIF* value) [17,18] and inter-correlation among the descriptors (*ICAP*). The P value of each substituent constant is less than 0.01, advocates linear relationship between the descriptors and activity. The *VIF* is defined as $1/(1 - r_i^2)$, where r_i is the multiple correlation coefficient for the i^{th} variable regressed on the $p - 1$ others, p being the variables contributed to the model. *VIF* value larger than 10 indicates that the information of the descriptor may be hidden by the correlation of the descriptors [19]. VIF is less than 1.2 for all the contributing descriptors revealed that the descriptors are fairly independent to each others. The low value of *ICAP* ($< \sim 0.240$) also supported fairly independent contribution. We have also made attempts to investigate predictive power of the proposed model by using quality factor (*QF*). This is calculated by Pogliani's method [20,21] which defines *QF* as the ratio of correlation coefficient to standard error of estimation (*SEE*). Larger value of *QF* (4.616) will be supporting for better predictive power of the model. For reliability of the model, we have calculated regression associated statistical parameter called probable error of correlation (*PE*) [22]. Goodness of fit calculated as $PE = 2(1 - r^2)/3\sqrt{n}$, if the value of correlation coefficient (r) is more than six time of *PE* than the expression is good and reliable. In this model, value of coefficient of correlation is significantly higher then 6*PE*, supporting reliability and goodness.

The model was further analyzed for the outlier by Z-score method (*Z-value*). Outlier test has been facilitating in identification of unexplainable structurally diverse analogs. The best QSAR model should not have any outlier. The *Z-value* for individual compounds lies within the specific range ($<|2.5|$), indicated absence of outliers.

The chance of fortuitous correlation were checked with the help of randomized biological activity test (*Chance*), 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, the value of *Chance* in model 1 is less than 0.001 suggested that the results were not based on prospective correlation.

To further access the robustness and statistical confidence of the model, bootstrapping analysis was performed. The bootstrapping analysis gives an overview about contribution of individual molecules to the QSAR model. The r_{bs}^2 is average squared correlation coefficient calculated during the validation procedure which is computed from a subset of compounds used one at a time for the validation procedure while S_{bs} is the standard deviation in multiple run of a given data set. If the value of r_{bs}^2 is at par to conventional r^2 and S_{bs} is low, than the model is robust and promising. In our study both values ($r_{bs}^2 = 0.875$ & $S_{bs} = 0.079$) fall within the agreement.

The quality of the final equation obtained via SEQ-MLR was confirmed by means of the Kubinyi function (*FIT*) [23,24]. It is closely relates to the Fisher ratio (F), although the main disadvantage of F is its sensitivity to changes in d. If d is small sensitivity is high and *vice- versa*. The *FIT* has a low sensitivity towards changes in d values, as long as they are small numbers, and a substantially increasing sensitivity for large d values. *FIT* defined as; $FIT = \{r^2 \cdot (n - = \{r^2 \cdot (n - d - 1)\}/\{(n + d^2)(1 - r^2)\}$, where n is number of compounds, d is optimal descriptors and r^2 is squared correlation coefficient. The model showed high value of *FIT* (5.016) revealed quality of fitness.

Predictivity of the model was assured with the help of cross-validated constraints like q^2, S_{PRESS} and S_{DEP} obtained by 'leave one out (loo)' method. In this, model was built with $N - 1$ compounds and the N^{th} compound is predicted (Table III & Figure 1). Each compound is left out from the model derivation and predicted in turn. The value of q^2 is the basic requirement for declaring a model to be a valid one is $q^2 > 0.5$ [25]. The consistency of the model supported by $q^2 = 0.828$, $S_{PRESS} = 0.291$ and $S_{DEP} = 0.252$ values.

Hansch approach suggested that molar refractivity and electronic property (Hammet's constant) at R_1 position play significant role in explaining the variance in activity similarly F_{hyd} at R position contributed significantly. Molar refractivity (MR_1) which is representative of molar volume and polarizability of the substituents play crucial role at R_1 position of the scaffold and suggested that bulkier group with optimum polarizability is favorable. Hammet's constant is an

Table III. Value of substituent constant and Hansch analysis data of calculated, calculated (loo), residual and *Z-score* of *N*-hydroxythiosemicarbazone analogs.

	Substituent Constant				Model-2 (pMIC)			
Comd. No.	MR_1	σ_1	F_{hyd}	Cal⋆	Cal_{res}†	*Z-score*	Cal(loo)‡	$Cal(loo)_{res}$¶
1	1.03	0.00	0.23	4.414	−0.221	−1.192	4.480	−0.287
2	2.85	0.12	0.23	4.598	−0.069	−0.372	4.609	−0.080
3	2.85	−0.37	0.23	4.208	0.321	1.733	4.102	0.427
4	15.35	−0.83	0.23	4.447	−0.166	−0.899	4.768	−0.488
5	7.87	−0.27	0.23	4.530	0.295	1.594	4.485	0.340
6	5.65	0.10	0.23	4.717	0.170	0.917	4.701	0.186
7	10.72	−0.25	0.23	4.684	−0.119	−0.642	4.714	−0.149
8	6.03	0.23	0.23	4.839	0.046	0.250	4.834	0.051
9	7.36	0.78	0.23	5.341	−0.153	−0.826	5.416	−0.229
10	7.36	0.71	0.23	5.285	−0.097	−0.525	5.325	−0.137
11	1.03	0.00	0.77	4.923	−0.098	−0.527	4.956	−0.130
12	2.85	0.12	0.77	5.106	−0.249	−1.345	5.149	−0.292
13	5.65	0.10	0.77	5.226	−0.070	−0.376	5.233	−0.077
14	6.03	0.23	0.77	5.348	0.147	0.796	5.333	0.161
15	7.36	0.78	0.77	5.849	0.259	1.397	5.754	0.354
16	8.88	0.23	1.90	6.549	0.003	0.018	6.540	0.013

⋆ Calculated data of the compounds using model; † Residual value of calculated data; ‡ Calculated (loo) data of the compounds using leave-one-out method; ¶ Residual value of calculated (loo) data.

electron-withdrawing property and is quite favorable to the activity, which indicates R_1 position play key role in the ionic interaction of the ligand with macromolecules. Hence the electronic interaction seems to be dominating for the activity of the compounds at R_1 position. F_{hyd} is Leo hydrophobic fragmental constant, which explains the interactions of the fragment/atoms with receptor. The positive contribution of F_{hyd} revealed that hydrophobic nature of substituent at R position is complimentary for the activity.

Structural requirements of *N*-Hydroxythiosemicarbazone analogs for anti-tuberculosis were further explored through topological indices. Regression of topological indices with inhibitory activity gave significant expression as Equation (5) (model-3).

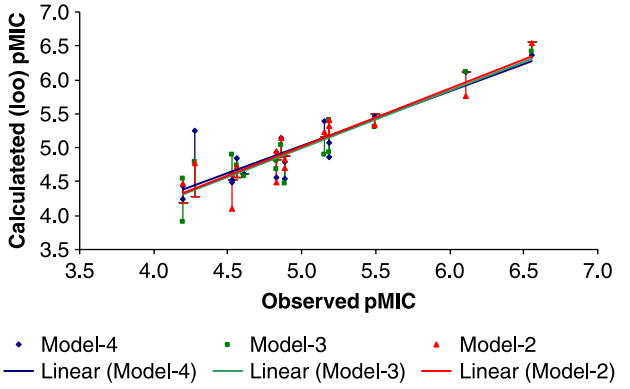


Figure 1. Plot of calculated (loo) MIC against observed MIC by models 2, 3 and 4.

Activity and statistical data showed in Tables IV and V respectively.

$$\begin{aligned} \text{pMIC} &= 0.042(\pm 0.004)D/D + 4.937(\pm 1.087) \\ &\quad \times SIC1 - 0.340(\pm 0.056)VRA2 + 0.181 \\ n &= 18,\ \ r = 0.951,\ \ r^2_{adj} = 0.885,\ \ SEE = 0.212, \\ F &= 44.575,\ \ F_{\max} = 29.308 \end{aligned} \tag{5}$$

Topological indices model showed correlation coefficient 0.951, which account for 88.5% explain variance in the activity (Figure 1). This model also explains the inhibitory activity of moderately diverse scaffolds as compare to Hansch substituent constant model (Equation (4)).

Topological indices model showed that *D/D* and *SIC1* contributed positively while *VRA2* contributed negatively. The D/D is distance/detour index, calculated as the half-sum of the entries of the distance/-detour quotient matrix. It was proposed as an index of molecular cyclicity, showing regular variation with increase in cyclicity in graphs of the same size. *SIC1* is Structural Information Content with neighborhood symmetry of 1-order. *SIC1* indices are calculated based on the pair wise equivalence atoms in a hydrogen-filled molecule. A pair of atoms are said to be equivalent at a particular level-r, if they are same to the element and their neighborhood is equivalent up to level-r. The positive contribution of equivalence neighborhoods pair of atoms *a* & *b* up to level-1 in structural architect suggested that this pharmacophoric feature is crucial for the activity. *VRA2* is average Randic-type eigenvector-based index from adjacency

Table IV. Calculated, calculated (loo), residual and *Z-score* of *N*-hydroxythiosemicarbazone analogs obtained from model-3 and model-4.

	Model-3 (pMIC)					Model-4 (pMIC)				
Comd. No.	Cal⋆	Cal_{res}†	*Z-score*	Cal(loo)‡	$Cal(loo)_{res}$¶	Cal⋆	Cal_{res}†	*Z-score*	Cal(loo)‡	$Cal(loo)_{res}$¶
1	4.457	−0.264	−1.373	4.549	−0.356	4.228	−0.034	−0.200	4.236	−0.042
2	4.608	−0.079	−0.410	4.617	−0.088	4.502	0.027	0.157	4.499	0.030
3	4.855	−0.326	−1.698	4.901	−0.372	4.522	0.007	0.043	4.521	0.008
4	4.506	−0.225	−1.174	4.794	−0.514	4.434	−0.153	−0.890	5.250	−0.970
5	4.816	0.009	0.047	4.816	0.010	4.588	0.237	1.374	4.568	0.258
6	4.529	0.358	1.863	4.480	0.407	4.577	0.310	1.799	4.538	0.349
7	4.697	−0.132	−0.687	4.736	−0.171	4.818	−0.253	−1.467	4.842	−0.277
8	4.890	−0.004	−0.022	4.890	−0.005	4.791	0.094	0.547	4.782	0.104
9	5.375	−0.187	−0.972	5.412	−0.224	4.889	0.299	1.735	4.863	0.325
10	4.958	0.230	1.199	4.924	0.264	5.080	0.108	0.624	5.069	0.118
11	4.707	0.118	0.614	4.678	0.147	4.915	−0.089	−0.519	4.928	−0.103
12	5.018	−0.161	−0.837	5.032	−0.174	5.120	−0.263	−1.523	5.147	−0.290
13	4.932	0.225	1.169	4.895	0.262	5.354	−0.198	−1.149	5.397	−0.241
14	5.331	0.163	0.851	5.302	0.192	5.469	0.026	0.153	5.464	0.031
15	6.111	−0.003	−0.016	6.112	−0.004	6.108	0.000	0.001	6.108	0.000
16	6.494	0.059	0.307	6.416	0.137	6.487	0.066	0.382	6.359	0.194
17	3.993	0.201	1.044	3.907	0.287	4.374	−0.180	−1.043	4.438	−0.245
18	4.591	0.018	0.095	4.582	0.028	4.614	−0.004	−0.025	4.616	−0.006

⋆ Calculated data of the compounds using model; † Residual value of calculated data; ‡ Calculated (loo) data of the compounds using leave-one-out method; ¶ Residual value of calculated (loo) data.

matrix. It can be related with molecular branching and accounting for contributions rooted in path clusters, clusters and chains of different lengths. The sound and robust topological indices model suggested that codify structural information contained in "molecular connectivity" could be explored for further development of *N*-Hydroxythiosemicarbazones.

In search of the conformational effect on QSAR study, the inhibitory activity was correlated with three dimensional descriptors of the energy minimized molecules. A tri-parametric expression (Equation (6)) was selected on the basis of statistical criteria. Plot, activity and statistical data showed in Figure 1, Tables IV and V respectively.

$$\begin{aligned} \text{pMIC} = {} & 1.106 + 0.191(\pm 0.015)G1 \\ & - 0.344(\pm 0.054)RDF115u \\ & + 1.817(\pm 0.202)Mor09u \quad n = 18, \\ r = 0.961, \;\; & r^2_{adj} = 0.907, \;\; SEE = 0.190, \\ F = 56.459, \;\; & F_{\max} = 54.092 \end{aligned} \tag{6}$$

Table V. Comparative statistics of QSAR models obtained from Hansch substituent constant (Model-2), topological descriptors (Model-3) and conformational dependent descriptors (Model-4).

Statistical Parameter	Model-2	Model-3	Model-4
n	16	18	18
r	0.955	0.951	0.961
r^2_{adj}	0.891	0.885	0.907
SEE	0.207	0.212	0.190
F	41.802	44.575	56.459
F_{max}	9.831	29.308	54.092
FIT	5.016	4.952	6.273
QF	4.616	4.495	5.059
PE	0.015	0.015	0.012
AIC	0.071	0.070	0.057
ICAP	<0.239	<0.434	<0.556
r^2_{bs}	0.875	0.895	0.929
S_{bs}	0.079	0.105	0.065
Chance	<0.001	<0.001	<0.001
q^2	0.828	0.832	0.760
S_{PRESS}	0.290	0.282	0.337
S_{DEP}	0.251	0.248	0.297
$q2_{(nout)}$⋆	–	0.821	0.781
$S_{PRESS(nout)}$	–	0.266	0.314
$S_{DEP(nout)}$	–	0.256	0.302
VIF of Descriptors	<1.111	<1.981	<2.051

⋆ In leave n out method, value of n is three.

Conformational dependent descriptors based Equation (model-4) contributed positively by gravitational index (*G1*) and 3D molecular representation of structure based on electron diffraction code (Mor09u), while radial distribution function (*RDF115u*) contributed negatively. *G1* is gravitational index [26]. It is a molecular descriptor reflecting the mass distribution in a molecule, defined as:

$$\text{GI} = \sum_{i+1}^{nAT-1} \sum_{j=i+1}^{nAT} m_i \cdot m_j / r_{ij}^2$$

Where, m_i and m_j are the atomic masses of the considered atoms, r_{ij} the corresponding interatomic distances, nAT the number of atoms of the molecule. The G1 index takes into account of all atom pairs in the molecule. This index is related to the bulk cohesiveness of the molecule accounting, simultaneously, for both atomic masses (volumes) and their distribution within the molecular space.

Mor09u [27–30] is 3D molecular representation of structure based on electron diffraction code (MoRSE code). MoRSE code was calculated by summing atom weights viewed by a different angular scattering function. The values of these code functions were calculated at 32 evenly distributed values of scattering angle(s) in the range of 0–31 $Å^{-1}$ from the three dimensional atomic co-ordinates of a molecule. The 3D-MoRSE code was calculated using following expression;

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}}$$

Where, s is scattering angle, r_{ij} is interatomic distance of i^{th} and j^{th} atom, A_i and A_j are atomic properties of i^{th} and j^{th} atom respectively including Van der Waals volume, atomic mass, Sanderson atomic electronegativity and atomic polarizability. The positively contribution of *Mor09u* revealed that sum of the properties calculated for the atoms (unweighted) from the three dimensional atomic co-ordinates of a molecule is decisive for explaining the enzyme ligand interaction.

RDF descriptors belong to the class of radial distribution function descriptors [31–33] are based on the distance distribution in the geometrical representation of the molecule. In addition to interatomic distances in the entire molecule, the RDF also provides valuable information about bond distances, ring types, planar and non-planar systems, atom types and other important structural motifs.

The RDF code has been proven to be a good representation for the 3D structure which has several merits like independence from the number of atoms; confidence regarding the three-dimensional arrangement of the atoms and invariance against translation and rotation of the entire molecule.

The RDF of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius *r*. The RDF used in this work is as follows:

$$g(r) = f \sum_{1}^{N-1} \sum_{j>i}^{N} AiAj e^{-B(r-r_i)^2}$$

$$f = 1 \Big/ \sqrt{\sum_{r} [g(r)]^2}$$

Where *f* is a scaling factor, *N* is the number of atoms, *A* are atomic properties of atoms *i* and *j*, *B* is smoothing parameter defines the probability distribution of the individual distances, r_{ij} is distance between the atoms *i* and *j*, $g(r)$ was calculated at a number of discrete points with defined intervals.

Each molecule was represented by a vector of length 32. The parameter *B* was set to 25 $Å^{-2}$ corresponding to a total resolution of 0.2 Å in the defined distance *r*.

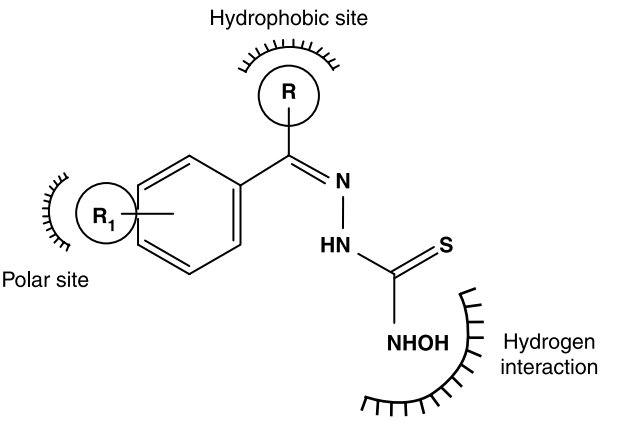


Figure 2. Illustrative representation of *N*-hydroxythiosemicarbazones interaction with receptor.

The RDF for the structure derivations was calculated with the atomic properties. *RDF115u* is radial distribution function at 11.5 Å interatomic distance un-weighted and contributed negatively.

We have shown that Fujita-Ban, Hansch substituent constant, topological descriptors and conformational dependent descriptors approaches are able to explain the anti-tubercular activity of *N*-Hydroxythiosemicarbazone analogs. Fujita-Ban & Hansch approaches having certain limitation, however, another approache showed their significant role in explaining activity of modified scaffolds. QSAR study of *N*-Hydroxythiosemicarbazone analogs with inhibitory activity furnished some important structural insights i.e. the R position is more prone for improving inhibitory activity and R_1 position play decisive role in the ionic interaction of the ligand with macromolecules. On the basis of the findings, interaction of *N*-hydroxythiosemicarbazones with macromolecule is illustrated in Figure 2.

Acknowledgements

The authors are grateful to the director of Shri G. S. Institute of Technology and Science, Indore for providing facilities for this work. The author R.A.G. is grateful to UGC, New Delhi, for providing junior research fellowship.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

[1] Ginsburg AS, Grosset JH, Bishai WR. Fluoroquinolones, tuberculosis, and resistance. Lancet Infect Dis 2003;3: 432–442.

[2] Havlir DV, Barnes PF. Tuberculosis in patients with human immunodeficiency virus infection. N Engl J Med 1999;340: 367–373.

[3] Sonnenberg P, Glynn JR, Fielding K, Murray J, Godfrey-Faussett P. How soon after infection with HIV does the risk of tuberculosis start to increase? A retrospective cohort study in South African gold miners. J Infect Dis 2005;191:150–158.
[4] World Health Organization. The World Health Organization Global Tuberculosis Program 2003, Web site: http://wwwwhoint/gtb
[5] Honer zu Bentrup K, Russell DG. Mycobacterial persistence: Adaptation to a changing environment. Trends Microbiol 2001;9:597–605.
[6] Janin YL. Antituberculosis drugs: Ten years of research. Bioorg Med Chem 2007;15:2479–2513.
[7] Sriram D, Yogeeswari P, Dakhla P, Senthil Kumar P, Banerjee D. N-hydroxythiosemicarbazones: Synthesis and *in-vitro* antitubercular activity. Bioorg Med Chem Lett 2007; 17:1888–1893.
[8] Hansch C, Leo A. Substituent constants for correlation analysis in chemistry and biology. New York: John Wiley; 1979.
[9] Wolff ME. Burger's medicinal chemistry and drug discovery. New York: John Wiley & Sons; 1994.
[10] Skagerberg B, Bonelli D, Clementi S, Cruciani G, Ebert C. Principal properties for aromatic substituents a multivariate approach for design in QSAR. Quant Struct Act Relat 1989;8: 32–38.
[11] CS Chem Office, Version 80. Cambridge soft corporation, Software publishers Association, 1730 M Street, suite 700, Washington DC 20036 202;: 452-1600, USA
[12] Todeschini R, Consonni V. DRAGON-software for the calculation of molecular descriptors, rel 112 for Windows, 2001.
[13] Gupta AK, Arockia Babu M, Kaskhedikar SG. VALSTAT: A program for quantitative structure activity relationship studies and their validations. Ind J Pharm Sci 2004;66:396–402.
[14] Schaper KJ. Free-Wilson-type analysis of non-additive substituent effects on THPB dopamine receptor affinity using artificial neural networks. Quant Struct Act Relat 1999;18: 354–360.
[15] Wold S, Eriksson L. Chemometric methods in molecular design. In: van de Waterbeemd H, editor. Weinheim: VCH; 1995. p 321.
[16] Livingstone DJ, Salt DW. Judging the significance of multiple linear regression models. J Med Chem 2005;48:661–663.
[17] Chatterjee S, Hadi A, Price B. Regression analysis by examples. 3rd ed. New York: Wiley-VCH; 2000.
[18] Shapiro S, Guggenheim B. Inhibition of oral bacteria by phenolic compounds part 1 QSAR analysis using molecular connectivity. Quant Struct Act Relat 1998;17:327–337.
[19] Cho DH, Lee SK, Kim BT, No KT. Quantitative structure-activity relationship (QSAR) study of new fluorovinyloxyacetamides. Bull Korean Chem Soc 2001;22:388–394.
[20] Pogliani L. Structure property relationships of amino acids and some dipeptides. Amino acids 1994;6:141–153.
[21] Pogliani L. Modeling with special descriptors derived from a medium-sized set of connectivity indices. J Phys Chem 1996; 100:18065–18077.
[22] Mandloi D, Joshi S, Khadikar PV, Khosla N. QSAR study on the antibacterial activity of some sulfa drugs: Building blockers of Mannich bases. Bioorg Med Chem Lett 2005;15:405–411.
[23] Kubinyi H. Variable selection in QSAR studies II: A highly efficient combination of systematic search and evolution. Quant Struct Act Relat 1994;13:393–401.
[24] Kubinyi H. Variable selection in QSAR studies I: An evolutionary algorithm. Quant Struct Act Relat 1994;13: 285–294.
[25] Golbraikh A, Tropsha A. Beware of q^2! J Mol Grap Mod 2002; 20:269–276.
[26] Katritzky AR, Mu L, Lobanov VS, Karelson M. Correlation of boiling points with molecular structure 1 a training set of 298 diverse organics and a test set of 9 simple inorganics. J Phys Chem 1996;100:10400–10407.
[27] Gasteiger J, Sadowski J, Schuur J, Selzer P, Steinhauer L, Steinhauer V. Chemical information in 3D space. J Chem Inf Comput Sci 1996;365:1030–1037.
[28] Schuur JH, Selzer P, Gasteiger J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. J Chem Inf Comput Sci 1996;362: 334–344.
[29] Schuur J, Gasteiger J. Infrared spectra simulation of substituted benzene derivatives on the basis of a novel 3D structure representation. J Anal Chem 1997;69:2398–2405.
[30] Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim: Wiley-VCH; 2000.
[31] Hemmer MC, Steinhauer V, Gasteiger J. The prediction of the 3D structure of organic molecules from their infrared spectra. J Vibrat Spectrosc 1999;19:151–164.
[32] Fedorowicz A, Zheng L, Singh H, Demchuk E. QSAR study of skin sensitization using local lymph node assay data. Int J Mol Sci 2004;5:56–66.
[33] Hemmer MC, Gasteiger J. Prediction of three-dimensional structure using information from infrared spectra. Anal Chim Acta 2000;420:145–154.