Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 1109–1116
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Rationalization of physicochemical characters of oxazolyl thiosemicarbazone analogs towards multi-drug resistant tuberculosis: A QSAR approach

Revathi A. Gupta, Arun K. Gupta, Love K. Soni, S.G. Kaskhedikar*

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

Received 23 September 2006; received in revised form 11 January 2007; accepted 12 January 2007
Available online 27 January 2007

Abstract

The emergence of multi-drug resistant (MDR) strains of *Mycobacterium tuberculosis* and the continuing pandemic of tuberculosis emphasizes the urgent need for the development of new and potent anti-tubercular agents. In an effort to develop new and more effective agents to treat tuberculosis emphasis was focused on quantification of structure–activity relationship of oxazolyl thiosemicarbazone derivatives. The *de novo* analysis gave insight to some important structural features i.e. nitro group on phenyl ring at R_1 position is optimal for the activity and might be responsible for electronic interaction, while phenyl ring at R position interact with the hydrophobic pocket more effectively as compared to unsubstituted or methyl substituted analogs. Hansch approach offered the understanding and parameterization of interactions of the inhibitor with receptor. Similarly QSAR analysis gave some important physicochemical properties, i.e. empirical aromatic index (ARR) and 3D-MoRSE code value of scattering angle at 8Å^{-1}. These two physicochemical properties shall be helpful in the development of more potent analogs.

Keywords: QSAR; Oxazolyl thiosemicarbazone analogs; Tuberculosis

1. Introduction

Tuberculosis (TB) is one of the oldest and most pervasive diseases in history. Worldwide, TB is caused by *Mycobacterium tuberculosis* (MTB), and to a lesser degree by *Mycobacterium bovis* and *Mycobacterium africanum*, and continues to be a major disease of global importance infecting at least one third, or two billion, of the world's population.

Several antibiotics for the treatment of TB were discovered and developed in the 1940s and 1950s and in the subsequent decades the incidence of the disease declined, particularly in developed countries [1]. This trend has been reversed in the past twenty years, and this reversal is thought to be driven by several factors, including poverty, overcrowding and the synergy between HIV and TB [2,3]. This synergy is particularly sinister, where active TB infection arises after immune suppression by HIV. Where there is co-infection with TB and HIV, the risk of death is twice that of a person infected with HIV alone [2].

No new antibiotics against TB have been developed in the past 30 years. There are three front line antibiotics, isoniazid, rifampin and pyrazinamide, and several second-tier antibiotics, including ethionamide, streptomycin and *para*-aminosalicylic acid.

To treat an infection, a cocktail of drugs including, for example, isoniazid, rifampin, ethambutol and pyrazinamide are prescribed for two months followed by a continuation phase in which isoniazid and rifampin are taken. Long-term therapies lasting between six and nine months have frequently led to patient non-compliance and, in turn, contributed to the emergence of multi-drug resistant TB (MDR-TB) [4]. MDR strains, such as the notorious strain W [5], are increasingly being found which are resistant to many first-line drugs including

* Corresponding author. Tel./fax: +91 0731 2546031.
E-mail address: arunkg_73@hotmail.com (S.G. Kaskhedikar).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.01.018

isoniazid, rifampin, ethambutol, and pyrazinamide, as well as some of the second-line drugs, such as ethionamide, cycloserine, thiacetazone and the quinolones.

The cost of treating a patient carrying MDR-TB is much greater, typically running into tens of thousands of dollars per patient, than for patients carrying a drug-sensitive strain [6]. Without effective treatments, the fear is that the number of infections caused by MDR-TB will increase out of control.

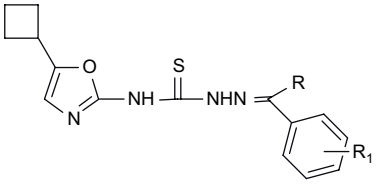
Despite of this, an urgent need for new antibiotics towards TB is widely acknowledged and a Global Alliance for TB Drug Development has been established with goals that include increasing compliance with current therapies, shortening treatment times with the development of new adjuvant drugs, improved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

The QSAR analysis of the anti-mycobacterial inhibitors is the recent interested area of research. The QSAR was performed on a series of oxazolyl thiosemicarbazone analogs [7]. The emphasis was focused on the quantification of structure–activity relationship with a view to delineate the influence of key mycobacterial inhibitory activity, which will aid in the designing of potent and safer inhibitors. The quantification of responsible physicochemical properties was done with the help of regression techniques.

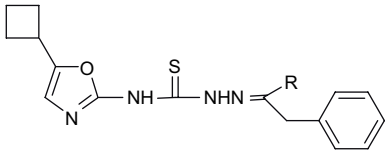
2. Experimental

The multi-drug resistant tuberculosis minimum inhibitory concentration data of oxazolyl thiosemicarbazone analogs was taken from the reported work of Sriram et al. [7] (Table 1). The inhibitory data (MIC in μg) 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

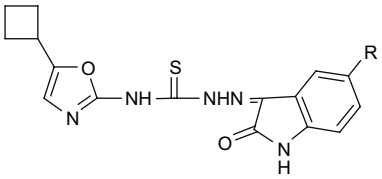
Table 1
Structure and activity of oxazolyl thiosemicarbazone analogs



Comp No. **1–17**



Comp. No. **18**



Comp No. **19 & 20**

Compound	R	R_1	MIC[a] (μg/ml)	pMIC (μM)[b]
1	H	2-OH	6.25	4.704
2	H	2-NO_2	0.78	5.646
3	H	3-NO_2	0.78	5.646
4	H	4-NO_2	0.39	5.947
5	H	4-CH_3	3.12	5.003
6	H	4-Cl	1.56	5.332
7	H	4-$N(CH_3)_2$	3.12	5.042
8	H	4-CH_3O	6.25	4.723
9	H	4-OH, 3-CH_30	12.5	4.443
10	CH_3	H	3.12	5.003
11	CH_3	2-OH	3.12	5.025
12	CH_3	4-OH	3.12	5.025
13	CH_3	4-CH_3	0.78	5.624
14	CH_3	4-NH_2	0.78	5.626
15	CH_3	4-NO_2	0.20	6.255
16	C_6H_5	H	0.10	6.576
17	C_6H_5	4-Br	0.05	6.959
18	$C_6H_5CH_2$	–	6.25	4.811
19	H	–	25.00	4.135
20	F	–	12.5	4.459

[a] Minimum inhibitory concentration against multi-drug resistant *Mycobacterium tuberculosis*.
[b] Negative logarithm of minimum inhibitory concentration in μM.

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, Hansch approach was carried out to establish correlations between MDR-TB inhibitory activity and various substituent constants at position R and R_1 of the molecule (Table 1). Values of substituent constants like hydrophobic (π), steric (Molar refractivity or MR), hydrogen acceptor (HA), hydrogen donor (HD), and electronic descriptor (field effect or $\mathscr{F}$, resonance effect or $\mathscr{R}$, and Hammet's constant or 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] version 8.0, and Dragon [12] program while the regression analysis was carried out on VALSTAT [13]. Structures of all the compounds were sketched using 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 variables 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) cross-validation method [14,15] considering predicted residual sum of square (PRESS), total sum of squares (SSY), cross-validated squared correlation coefficient (q^2), and standard deviation error of prediction (S_{DEP}). The q^2 is defined as

$$q^2 = 1 - \sum (Y_{pred} - Y_{act})^2 / \sum (Y_{act} - Y_{mean})^2$$

where Y_{pred}, Y_{act}, and Y_{mean} are predicted, actual and mean values of the pMIC, respectively. $\sum (Y_{pred} - Y_{act})^2$ is the predictive residual error sum of squares (PRESS). PRESS is an important cross-validation parameter as it is a good approximation of the real predictive error of the models. Bootstrapping analysis was performed to ascertained robustness and statistical confidence against the model. Bootstrapping squared correlation coefficient (r^2_{bs}) is the average squared correlation coefficient of subset of compounds used in regression. Chances of fortuitous correlation were tested with the help of *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). Test for outliers, which confirm the applicability of QSAR equation on the structural analogs were performed with the help of Z-score value. The external predictive power of the equation has been analyzed with the help of test set using predictive correlation coefficient (r^2_{pred}).

3. Results and discussion

Fujita-Ban analysis was carried out in order to find out *de novo* contribution of the substituents to the activity of the scaffold (Table 2). The multi-variant regression expression (Eq. (1)) indicated that many substituents have poor contribution to the inhibitory activity, which is further supported by high standard error of the substituent coefficient.

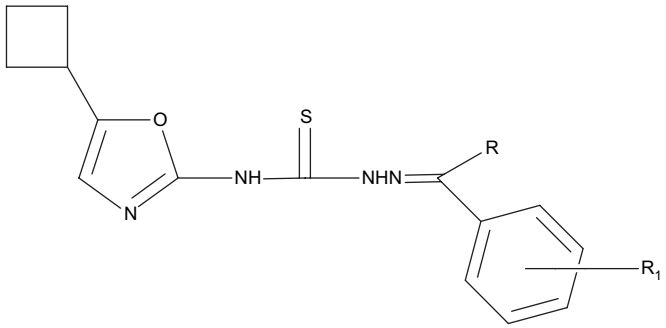
$$\begin{aligned}\text{pMIC} = {} & 0.416(\pm 0.102)\text{R-CH}_3 + 1.987(\pm 0.205)\text{R-C}_6\text{H}_5 + 0.069(\pm 0.162)\text{R}_1\text{-2OH} + 1.059(\pm 0.205)\text{R}_1\text{-2NO}_2 \\ & + 1.059(\pm 0.205)\text{R}_1\text{-3NO}_2 - 0.166(\pm 0.205)\text{R}_1\text{-3CH}_3\text{O} + 1.306(\pm 0.162)\text{R}_1\text{-4NO}_2 + 0.519(\pm 0.162)\text{R}_1\text{-4CH}_3 \\ & + 0.745(\pm 0.205)\text{R}_1\text{-4Cl} + 0.455(\pm 0.205)\text{R}_1\text{-4N(CH}_3)_2 + 0.136(\pm 0.205)\text{R}_1\text{-4CH}_3\text{O} + 0.022(\pm 0.177)\text{R}_1\text{-4OH} \\ & + 0.622(\pm 0.177)\text{R}_1\text{-4NH}_2 + 0.384(\pm 0.177)\text{R}_1\text{-4Br} + 4.587 \\ & n = 17,\ r = 0.998,\ r^2 = 0.996,\ \text{SEE} = 0.125,\ F = 34.434 \end{aligned} \quad (1)$$

Eq. (1) was further optimized by removing the substituents, which was contributed insignificant at 95% confidence level. Statistical significant trivalent expression (Eq. (2)) was considered.

$$\begin{aligned}\text{pMIC} = {} & 1.568(\pm 0.262)\text{R-C}_6\text{H}_5 - 0.757(\pm 0.358) \\ & \times \text{R}_1\text{-3CH}_3\text{O} + 0.901(\pm 0.262)\text{R}_1\text{-4NO}_2 + 5.200 \\ & n = 17, r = 0.894, r^2 = 0.799, \text{SEE} = 0.344, F = 17.181 \end{aligned} \quad (2)$$

Fujita-Ban analysis of MDR-TB inhibitory activity of oxazolyl thiosemicarbazone analogs inferred that the substitutions of electron withdrawing group at 3rd and 4th position of

Table 2
Fujita-Ban matrix of oxazolyl thiosemicarbazone analogs with calculated (cal) value



Compound	μ	R		R_1												Cal pMIC
		CH_3	C_6H_5	2-OH	2-NO_2	3-NO_2	3-CH_3O	4-NO_2	4-CH_3	4-Cl	4-$N(CH_3)_2$	4-CH_3O	4-OH	4-NH_2	4-Br	
1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5.200
2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	5.200
3	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5.200
4	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	6.101
5	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	5.200
6	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5.200
7	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5.200
8	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5.200
9	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	4.443
10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5.200
11	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	5.200
12	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	5.200
13	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	5.200
14	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	5.200
15	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6.101
16	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	6.768
17	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	6.768

phenyl ring are favorable. Especially substitution of nitro moiety at phenyl ring result in potent inhibition as compared to other substituted analogs. *De novo* techniques also suggested that the presence of another phenyl ring on carbimino terminal at R position causes increase in the inhibitory activity.

Fujita-Ban expression gave insight to some important structural features i.e. nitro group at R_1 position is optimal for the activity and might be responsible for electronic interaction. Presence of the phenyl ring on carbimino terminal (R position) might be helpful in accommodating the ligand in hydrophobic pocket more effectively as compared to unsubstituted or methyl substituted analogs.

Hansch approach was further employed in order to develop correlation between inhibitory activity and substituent constant using stepwise multiple linear regression method. Regression furnished several equations, but statistically significant trivariant equation was considered (Eq. (3)).

$$\begin{aligned}\mathrm{pMIC} &= 0.066\left(\pm 0.0064\right)\mathrm{MR} + 0.509\left(\pm 0.103\right)\mathscr{R}_1 \\ &\quad + 0.741\left(\pm 0.111\right)\sigma_1 + 5.056 \\ &\quad n = 17,\ r = 0.968,\ r^2 = 0.938, \\ &\quad \mathrm{SEE} = 0.191,\ F = 65.133 \qquad (3)\end{aligned}$$

Eq. (3) showed better correlation coefficient ($r = 0.968$), which accounted for 93.8% of the variance in the activity. The data showed overall internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(3,13\ \alpha\ 0.001)} = 11.9$. The P value of each substituent constant is less than 0.001, suggests linear relationship between the descriptors and activity. The cross-validated squared correlation coefficient ($q^2 = 0.888$), predictive residual sum of square ($S_{\mathrm{PRESS}} = 0.257$) and standard error of prediction ($S_{\mathrm{DEP}} = 0.225$) suggested a good internal consistency as well as predictive ability of the biological activity (Table 3). The bootstrapping correlation coefficient ($r^2_{\mathrm{bs}} = 0.913$) and randomized biological activity test (*Chance* < 0.001) are significant. The inter-correlation among the parameters is less than 0.150. Hansch approach suggested that steric effect (molar refractivity) at R position and electronic properties (resonance effect and Hammet's constant) at R_1 position contributed positively.

Molar refractivity (MR) [16,17] which is representative of molar volume and polarizability of the substituents play crucial role at R position of the scaffold and suggested that bulkier group with optimum polarizability is favorable. The substituted group might be interacting significantly at hydrophobic area of the active site. Hammet's constant is indicative of an electron-withdrawing property and it's contributed favorably to the inhibitory activity. The contribution of Hammet's constant at R_1 position suggested the role of substituent in the ionic interaction of the ligand with macromolecules. Thus the electronic interactions seem to be dominating for the activity of the compounds. However, the presence of

Table 3
Value of substituent constant and Hansch analysis data like calculated, calculated (loo), residual and Z-score of oxazolyl thiosemicarbazone analogs

Compound	Substituent constant			Hansch model (pMIC)				
	MR	π_1	σ_1	Cal[a]	Cal_{res}[b]	Z-score	Cal(loo)[c]	$Cal(loo)_{res}$[d]
1	1.03	−0.64	0.12	4.886	−1.057	−0.182	4.918	−0.214
2	1.03	0.16	0.71	5.731	−0.493	−0.085	5.752	−0.106
3	1.03	0.16	0.71	5.731	−0.493	−0.085	5.752	−0.106
4	1.03	0.16	0.78	5.783	0.953	0.164	5.735	0.213
5	1.03	−0.13	0.10	5.131	−0.742	−0.128	5.142	−0.139
6	1.03	−0.15	0.23	5.218	0.663	0.114	5.207	0.124
7	1.03	0.92	−0.83	4.977	0.379	0.065	4.682	0.360
8	1.03	−0.51	−0.27	4.664	0.346	0.060	4.652	0.071
9	1.03	−1.15	−0.25	4.353	0.523	0.090	4.296	0.147
10	5.65	0.00	0.00	5.427	−2.460	−0.424	5.460	−0.457
11	5.65	−0.64	0.12	5.190	−0.960	−0.165	5.215	−0.190
12	5.65	−0.64	−0.37	4.827	1.148	0.198	4.780	0.245
13	5.65	−0.13	0.10	5.435	1.098	0.189	5.423	0.201
14	5.65	0.18	0.12	5.608	0.103	0.018	5.606	0.020
15	5.65	0.16	0.78	6.087	0.973	0.168	6.043	0.212
16	25.36	0.00	0.00	6.724	−0.860	−0.148	6.857	−0.281
17	25.36	−0.17	0.23	6.808	0.879	0.151	6.673	0.287

[a] Calculated data of the compounds using model.
[b] Residual value of calculated data.
[c] Calculated (loo) data of the compounds using leave-one-out method.
[d] Residual value of calculated (loo) data.

hydrophobic pockets in the receptor might help in engulfing hydrophobic moieties present in the compounds. A general model of interaction of oxazolyl thiosemicarbazones ligand can be visualized as represented by Fig. 1.

Followed by Hansch approach series was subjected to QSAR analysis employing physiochemical properties. The series was divided into a training set of 15 compounds including compounds **1**, **3**, **4**, **6**, **7**, **9**–**14**, **16**, **17**, **19** and **20** (Table 1) and a test set of 5 compounds including compounds **2**, **3**, **8**, **15** and **18** (Table 1), on the basis of structural diversity and complete range of variation in inhibitory activity. The training set was subjected to sequential multiple linear regression analysis in order to a establish correlation between physicochemical parameters and inhibitory activity. Several significant equations with coefficient of correlation $(r) \geq 0.915$ were obtained, which accounts for more than 83.0% of the variance in the activity data (Eqs. (4–7)).

$$\mathrm{pMIC} = 0.968(\pm 0.105)\mathrm{MLOGP} - 0.987(\pm 0.421)\mathrm{Mor25m} + 3.298 \quad n = 15,\ r = 0.936,\ r^2 = 0.876,\ \mathrm{SEE} = 0.298,\ F = 42.424 \tag{4}$$

$$\mathrm{pMIC} = 1.095(\pm 0.287)\mathrm{Mor09u} + 11.506(\pm 1.475)\mathrm{ARR} + 1.730 \quad n = 15,\ r = 0.929,\ r^2 = 0.864,\ \mathrm{SEE} = 0.312,\ F = 38.095 \tag{5}$$

$$\mathrm{pMIC} = 0.419(\pm 0.218)\mathrm{MAXDN} + 0.970(\pm 0.113)\mathrm{MLOGP} + 2.060 \quad n = 15,\ r = 0.928,\ r^2 = 0.862,\ \mathrm{SEE} = 0.315,\ F = 37.427 \tag{6}$$

$$\mathrm{pMIC} = 1.058(\pm 0.330)\mathrm{Mor09m} + 12.716(\pm 1.649)\mathrm{ARR} + 1.936 \quad n = 15,\ r = 0.915,\ r^2 = 0.838,\ \mathrm{SEE} = 0.341,\ F = 30.977 \tag{7}$$

A high correlation coefficient merely is not enough to select the equation as a model. Equations were screened through various internal and external statistical validation techniques. Internal statistical significance level of the equations was confirmed using sequential Fischer test, all the equations have significance level more than 99.9% as it exceeded the tabulated $F_{(2,12\ \alpha\ 0.001)} = 15.7$. Sequential Fischer test recommended that equations are applicable for more than 999 times out of 1000. The inter dependency of physicochemical properties for each equation was checked in order to confirm inimitable contribution of the properties to the expression. All the regression expressions were checked for the presence of

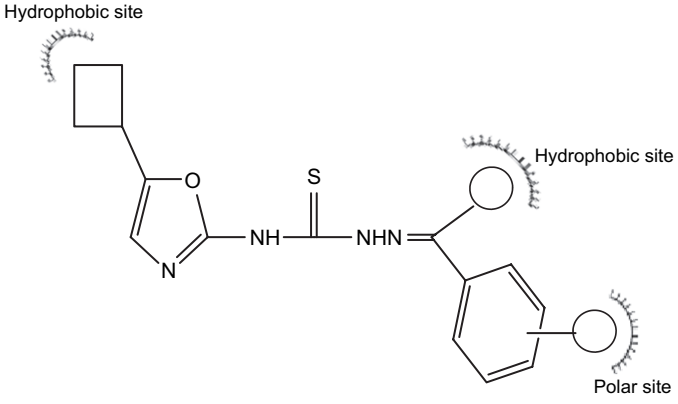


Fig. 1. Hypothetical model of interaction of oxazolyl thiosemicarbazones ligand with receptor.

Table 4
QSAR statistics of significant equations

Equation No.	n	r^2	SEE	F	ICAP[a]	r^2_{bs}	s_{bs}	*Chance*	q^2	S_{PRESS}	S_{DEP}	r^2_{pred}	Outlier
3	17	0.938	0.191	65.13	<0.143	0.913	0.073	<0.001	0.888	0.257	0.224	–	Nil
4	15	0.876	0.298	42.424	<0.320	0.880	0.083	<0.001	0.805	0.374	0.334	−0.951	Nil
5	15	0.864	0.312	38.095	<0.013	0.884	0.066	<0.001	0.731	0.439	0.393	0.599	Nil
6	15	0.862	0.315	37.427	<0.370	0.831	0.152	<0.001	0.787	0.390	0.349	−0.702	1
7	15	0.838	0.341	30.977	<0.215	0.815	0.146	<0.001	0.758	0.416	0.373	0.782	Nil

[a] The maximum limit of inter-correlation among the descriptors used in generation of equations.

outliers using *Z*-score method. This test confirmed the applicability of equation on structurally diverse analogs (Table 4). In the case of Eq. (6), one outlier was present. The presence of outlier revealed that physiochemical properties involved in Eq. (6) are not factual representatives for prediction of structurally diverse analog. Bootstrapping technique was employed to confirm the contribution of physicochemical properties of the molecules to the activity whether equi-intense or of different rank. The value of the bootstrapping squared correlation coefficient and the bootstrapping standard deviation implies that the equations were proper representatives of the group of analogs. The chance of fortuitous correlation was checked with the help of randomized biological activity test, the value of chance statistics (*Chance*) is less than 0.001. Data of chance statistics revealed that the results were not based on chance correlation. The internal consistency of the training set was confirmed through leave-one-out method of cross-validation. Although equations showed good internal consistency ($q^2 = 0.731-0.805$), they may not be applicable for the analogs, which were never used in the generation of the correlation. Therefore, the predictive power of Eqs. (4–7) was further confirmed by a test set of five compounds. Eqs. (5 and 7) showed r^2_{pred} (≥ 0.5) values, which revealed robustness and wide applicability of these equations. The statistical interpretation suggested that Eqs. (5 and 7) fulfills the statistical validation criteria to the significant extent and therefore would be considered as models for proposing more active compounds.

Eq. (5) (model-1) has a better correlation coefficient ($r = 0.929$), which accounted for more than 86.0% of the variance in the activity. The equation shows that in the multivariant model, the dependent variable can be predicted from a linear combination of the independent variables. The *P* value is less than 0.001 for each physicochemical parameter involved in model generation suggested statistical significant relationship between the descriptor and activity. The data showed overall internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(2,12\ \alpha\ 0.001)} = 15.7$. The cross-validated squared correlation coefficient ($q^2 = 0.731$), predictive

Table 5
Observed, calculated and calculated (loo) and predicted pMIC values with *Z*-score and residual of oxazolyl thiosemicarbazone analogs used in QSAR analysis

Compound	Model-1						Model-2				
	Obs[a]	Cal[b]	Cal_{res}[c]	*Z*-score	Cal(loo)/pred[d]	$Cal(loo)_{res}/pred_{res}$[e]	Cal[b]	Cal_{res}[c]	*Z*-score	Cal(loo)/pred[d]	$Cal(loo)_{res}/pred_{res}$[e]
1	4.704	5.100	−0.396	−1.371	5.131	−0.427	5.058	−0.354	−1.122	5.088	−0.384
3	5.646	5.500	0.146	0.506	5.482	0.164	5.498	0.148	0.468	5.480	0.166
4	5.947	5.760	0.188	0.650	5.740	0.207	5.847	0.100	0.316	5.834	0.113
6	5.332	5.292	0.040	0.137	5.288	0.043	5.710	−0.378	−1.198	5.903	−0.572
7	5.042	4.816	0.226	0.783	4.786	0.255	4.849	0.193	0.610	4.824	0.218
9	4.443	4.832	−0.389	−1.347	4.883	−0.440	4.535	−0.092	−0.293	4.553	−0.110
10	5.003	5.410	−0.407	−1.409	5.466	−0.463	5.431	−0.427	−1.354	5.500	−0.496
11	5.025	4.928	0.097	0.336	4.919	0.106	4.942	0.083	0.264	4.934	0.091
12	5.025	5.100	−0.075	−0.259	5.109	−0.084	5.071	−0.046	−0.145	5.076	−0.051
13	5.624	5.334	0.290	1.004	5.265	0.360	5.266	0.358	1.135	5.192	0.433
14	5.626	5.384	0.242	0.838	5.315	0.311	4.952	0.674	2.134	4.887	0.738
16	6.576	6.986	−0.410	−1.420	7.380	−0.804	6.657	−0.081	−0.257	6.746	−0.171
17	6.959	6.545	0.414	1.433	6.288	0.672	6.799	0.160	0.507	6.697	0.262
19	4.135	4.327	−0.192	−0.664	4.470	−0.335	4.609	−0.473	−1.500	4.748	−0.613
20	4.459	4.233	0.226	0.782	4.093	0.365	4.322	0.137	0.434	4.249	0.210
2	5.646	–	–	–	5.266	0.381	–	–	–	5.443	0.203
5	5.003	–	–	–	5.507	−0.503	–	–	–	5.341	−0.337
8	4.723	–	–	–	5.064	−0.340	–	–	–	5.121	−0.398
15	6.255	–	–	–	5.974	0.280	–	–	–	6.037	0.217
18	4.811	–	–	–	5.103	−0.292	–	–	–	4.903	−0.092

[a] Observed data of the compounds used in generation of model.
[b] Calculated data of the compounds using model.
[c] Residual value of calculated data.
[d] Calculated (loo) data of the compounds using leave-one-out method and predicted data of test set compound.
[e] Residual value of calculated (loo) data and predicted data compound.

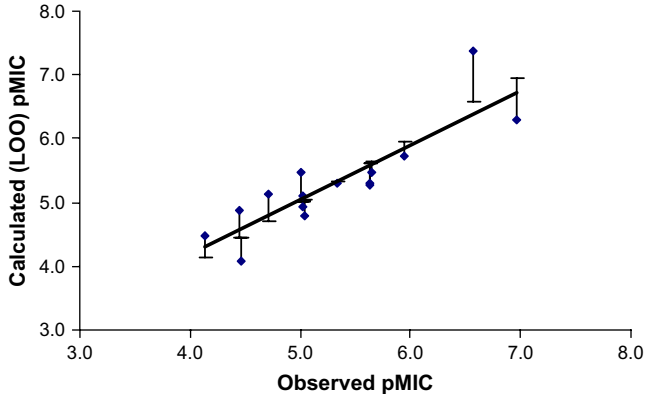


Fig. 2. Graphical representation of observed and calculated (loo) pMIC using Model-1.

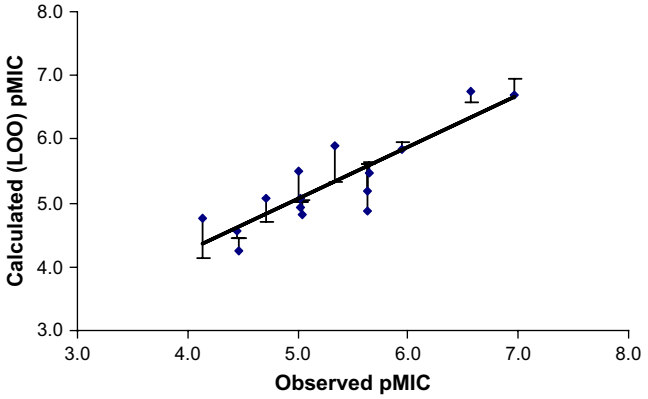


Fig. 4. Graphical representation of observed and calculated (loo) pMIC using Model-2.

residual sum of square ($S_{\mathrm{PRESS}} = 0.439$) and standard error of prediction ($S_{\mathrm{DEP}} = 0.392$) suggested a good internal consistency as well as predictive ability of the biological activity (Table 5 and Fig. 2). The r^2_{bs} is at par with the conventional squared correlation coefficient (r^2). Randomized biological activity test (*Chance* < 0.001) revealed that the results were not based on chance correlation. In chance statistics value of randomized r^2 is positioned between 0.000 and 0.602 and mean randomized r^2 value is 0.145 with standard deviation $\pm$ 0.125. The inter-correlation among the parameters is less than 0.320. Test data set gave significant predictive correlation coefficient ($r^2_{\mathrm{pred}} = 0.599$) with mild standard error of predictivity ($\mathrm{SEE}_{\mathrm{pred}} = 0.339$) (Tables 4 and 5 and Fig. 3).

Model suggested that Mor09u [18–21] and ARR [21] contributed positively to the inhibitory activity of oxazolyl thiosemicarbazones. Mor09u is 3D molecular representation of structure based on electron diffraction code (3D-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 $\mathrm{\AA}^{-1}$ from the three dimensional atomic coordinates 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 ith and jth atom, A_i and A_j are atomic properties of ith and jth atom, respectively, including atomic number, atomic mass, partial atomic charges, residual electro-negativities, and atom polarizability. ARR is an empirical aromatic index obtained as a function of the ratio between number of aromatic bonds and the total bonds in the H-depleted molecule.

Eq. (7) (model-2), showed correlation coefficient at par with model-1, similarly internal statistical significance level better than 99.9% and showed pretty good internal predictivity as compared to model-1 (Table 5 and Fig. 4). The external predictivity (0.782) improved significantly as compared (Tables 4 and 5 and Fig. 5) to 0.599 and low predictive standard error of estimation ($\mathrm{SEE}_{\mathrm{pred}} = 0.185$). In model-2 Mor09m [16–19] and ARR contributed positively to the inhibitory activity. Mor09m, 3D-MoRSE code was calculated by summing atom weights viewed by a different angular scattering function. In MoRSE code, *m* notation after digital value used for the atom weights was contributed by especially through atomic masses.

The quantification of the structural features of oxazolyl thiosemicarbazone analogs with inhibitory activities furnished

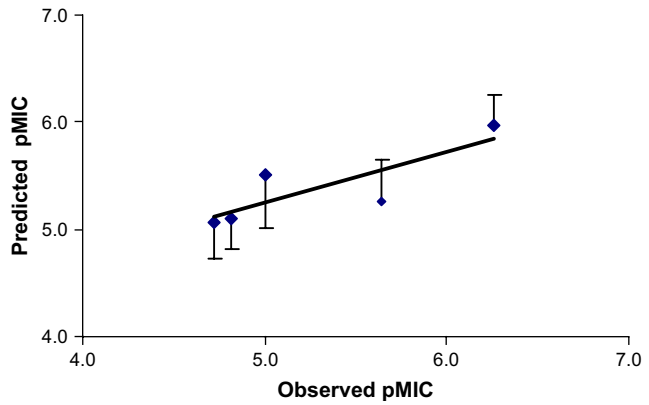


Fig. 3. Graphical representation of observed and predicted pMIC using Model-1.

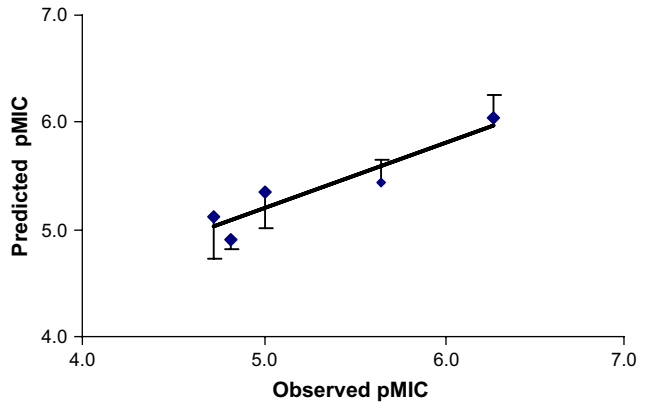


Fig. 5. Graphical representation of observed and predicted pMIC using Model-2.

some important structural insights i.e. empirical aromatic index (ARR) contributed positively to the activity. 3D molecular representation of structure based on electron diffraction (3D-MoRSE) code is the dominant structure feature, which is decisive in explaining the inhibitory activity. These structure features may be helpful in development of more safer and potent inhibitors.

References

[1] V.L. Arcus, J.S. Lott, M.J. Johnston, N.E. Baker, Drug Discov. Today 11 (2006) 28–34.
[2] D.V. Havlir, P.F. Barnes, N. Engl. J. Med. 340 (1999) 367–373.
[3] P. Sonnenberg, J.R. Glynn, K. Fielding, J. Murray, P. Godfrey-Faussett, S. Shearer, J. Infect. Dis. 191 (2005) 150–158.
[4] K.H.Z. Bentrup, D.G. Russell, Trends Microbiol. 9 (2001) 597–605.
[5] T.B. Agerton, S.E. Valway, R.J. Blinkhorn, K.L. Shilkret, R. Reves, W.W. Schluter, B. Gore, C.J. Pozsik, B.B. Plikaytis, C. Woodley, I.M. Onorato, Clin. Infect. Dis. 29 (1999) 85–92.
[6] C. Dye, B.G. Williams, M.A. Espinal, M.C. Raviglione, Science 295 (2002) 2042–2046.
[7] D. Sriram, P. Yogeeswari, R. Thirumurugan, R.K. Bhavana, J. Med. Chem. 49 (2006) 3448–3450.
[8] C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley, New York, 1979.
[9] M.E. Wolff, Burger's Medicinal Chemistry and Drug Discovery, John Wiley & Sons, New York, 1994.
[10] B. Skagerberg, D. Bonelli, S. Clementi, G. Cruciani, C. Ebert, Quant. Struct. Act. Relat. 8 (1989) 32–38.
[11] CS Chem Office, Version 8.0, Cambridge soft corporation, Software publishers Association, 1730 M Street, suite 700, Washington D.C. 20036, U.S.A., (202) 452–1600.
[12] R. Todeschini, V. Consonni, DRAGON-Software for the Calculation of Molecular Descriptors, rel. 1.12 for Windows (2001).
[13] A.K. Gupta, M.A. Babu, S.G. Kaskhedikar, Indian J. Pharm. Sci. 66 (2004) 396–402.
[14] K.J. Schaper, Quant. Struct. Act. Relat. 18 (1999) 354–360.
[15] S. Wold, L. Eriksson, in: H. van de Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH, Weinheim, 1995.
[16] M.H. Abraham, Chem. Soc. Rev. 22 (1993) 73–83.
[17] M.H. Abraham, A.J.M. Al-Hussaini, J. Environ. Monit. 3 (2001) 377–381.
[18] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, J. Chem. Inf. Comput. Sci. 36 (1996) 1030–1037.
[19] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 334–344.
[20] J. Schuur, J. Gasteiger, Anal. Chem. 69 (1997) 2398–2405.
[21] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.