Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

ISSN No: 0975-7384
CODEN(USA): JCPRC5

J. Chem. Pharm. Res., 2011, 3(2):479-488

Pyrimidinedione: Pharmacophore Optimization of Selective Thymidine Monophosphate Kinase inhibitors using Group QSAR Studies as Potential Antitubercular agents

Chitre T. S. and *Bothara K. G.

Dept .of Pharm. Chemistry (PG), AISSMS College of Pharmacy, Pune, Maharashtra, India

ABSTRACT

Tuberculosis (TB) is the second major cause of death from a single infectious agent among adults in developing countries, followed by HIV. The emergence of multi-drug resistant strains of Mycobacterium tuberculosis and revival of TB in the industrialized world due to HIV infections has rendered the quest for new drugs against TB a priority. In this work an effort is made to optimize the pharmacophore required for potent and selective inhibition of one essential enzyme of nucleotide metabolism, viz., thymidine monophosphate kinase (TMPKmt), by selecting reported series of TMPKmt inhibitors. A new Group-Based QSAR (G-QSAR) method which uses descriptors evaluated for the fragments of the molecules, generated using specific fragmentation rules for the selected dataset was carried out . G-QSAR was specifically done for knowing the structure activity relationship for carrying out variations in substitution at specific substitution sites. Also mathematical model for the prediction of activities of the new molecules was developed. G-QSAR studies were carried out using VLife Molecular Design Suite (V Life MDS) software.

Key Words: Pharmacophore Optimization, Thymidine monophosphate kinase of *M. tuberculosis* (TMPKmt), Simulating Annealing (SA), Genetic Algorithm (GA), G-QSAR.

INTRODUCTION

Mycobacterium tuberculosis is responsible for at least 2 million deaths globally per year. Due to demographic factors, socioeconomic trends, neglected tuberculosis control in many countries and HIV infection, this epidemic has been able to adopt such a huge proportion. [1] A peculiar aspect of its pathogenicity comes from the fact that it can remain quiescent and become active decades

later. [2] The current treatment of active TB includes a dosage regime of four drugs (Isoniazide, Rifampicin, Pyrazinamide, and Ethambutol) for at least six months. As a consequence of the prolonged duration, irregular treatment and the highly adaptive nature of organism to the surroundings, multidrug resistant (MDR) strain of *M. tuberculosis* have emerged. Therefore there is urgent need to cope with the current crisis. Since the determination of the *M. tuberculosis* genome sequence, this genomic information has been used to identify and validate targets as the basis for the development of new anti-tubercular agents. [3]

M. tuberculosis thymidine monophosphate kinase ($TMPK_{mt)}$ belongs to a large super family of nucleoside monophosphate kinases ($NMPK_s$). It catalyses the phosphorylation of deoxythymidine monophosphate (dTMP) to deoxytymidine diphosphate (dTDP) utilizing ATP as a phosphoryl donor. This step lies at the junction of the de novo and salvage pathway of thymidine triphosphate (TTP) metabolism and is the last specific enzyme for its synthesis. Also, the sequence of $TMPK_{mt}$ when compared with that of its human isozyme shows only 22% sequence identity. These characteristics make $TMPK_{mt}$ one of the potential targets for the design of new antitubercular drugs. Several dTMP derivatives were synthesized and studied for their effect on the $TMPK_{mt}$ [4].

Moreover, 3'-azido-3'-deoxythymidine monophosphate (AZTMP) was identified as a competitive inhibitor of $TMPK_{mt}$ with a K_i of 10 µM [4]. $TMPK_{mt}$ represents the first reported TMPK that does not phosphorylate AZTMP, a feature that could be exploited in the search for other selective inhibitors of $TMPK_{mt}$. Lavie et al. postulated that, in yeast TMPK, the azido groups interacts via its terminal nitrogen with the side –chain carboxyl of Asp^9, resulting in a P-loop displacement. Thus, Arg_{15} binds ATP less efficiently, slowing down catalysis. A similar mechanism may account for the lack of phosphorylation of AZTMP by $TMPK_{mt}$. Li de la Sterra et al. suggested that the azido group displaces the magnesium ion in the active site of $TMPK_{mt}$ (responsible for positioning one oxygen of the phosphate and Asp9), thereby abolishing catalysis. [1]

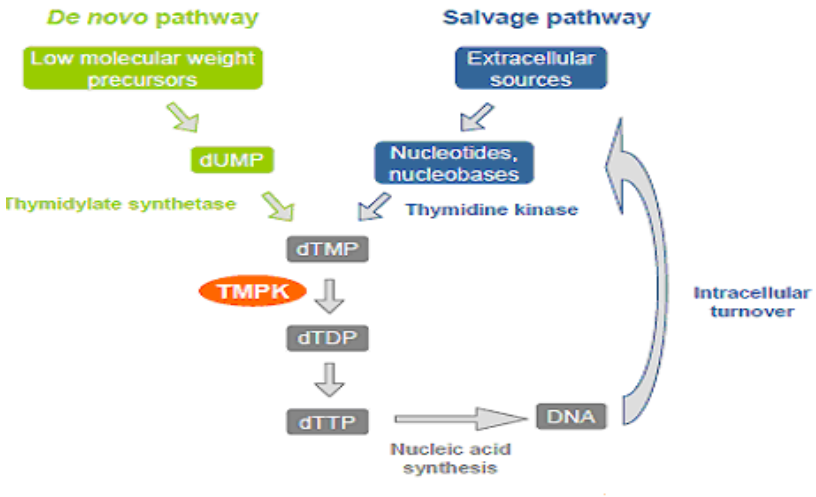


Fig. 1

Some thymidine 5'-O- monophosphate analogues were shown to be inhibitors of TMPKmt. [4, 5, 6] For therapeutic application these analogues should be administered as nucleosides, which

should be phosphorylated to their nucleotide monophosphate counterparts by *M. tuberculosis* .However it is reported that there is no Thymidine kinase (TK) activity in this bacterium[7,8] which seemingly renders the use of $TMPK_{mt}$ as a invalid target. Finding also have been reported that 5-bromo- 2'deoxyuridine (5 BrdU) and 3'-azido-3'- deoxythymidine (AZT) are inhibitors of $TMPK_{mt}$ as potent as their nonphosphorylated 5'-modified derivatives which means that no metabolic processing would be necessary for these compounds to become pharmacologically active, which has opened new avenues in the search for specific inhibitors of $TMPK_{mt}$ as antitubercular drugs.[9] Literature survey indicates that, in order to discover potent and selective inhibitors of TMPKmt, a fragment based de novo drug design program (LEA3D) have been applied to the dTMP binding site of TMPKmt with the aim of generating new ligand families.[10,11]

As compared to conventional QSAR methods (2D and 3D), a new Group-QSAR (G-QSAR) [12], method reported by Ajmani *et. al*. allows flexibility to the study of molecular substitution sites of interest. Also the method is independent of conformational analysis and alignment of the molecules, which is a necessary requirement of 3D QSAR. It is reported to provide clues about the substitution required at particular site as opposed to earlier QSAR models which do not clearly specify the site at which modification is required. The statistical results of G-QSAR are also reported to be easy to interpret as compared to conventional QSAR methods.

In continuation of our work in antimycobacterial area [13] and in an effort to optimize the pharmacophore required to selectively inhibit Thymidine monophosphate kinase of *M. tuberculosis* (TMPKmt), we have carried out G-QSAR to provide us with hints for the sites of improvement in the molecules. We have selected a reported series of substituted benzyl pyrimidines as inhibitors of TMPKmt [11] for the G-QSAR work.

EXPERIMENTAL SECTION

Biological data

Twenty five molecules, reported for their anti-tubercular effect [11] were selected for the present study. The structures of the compounds and their anti-tubercular activity data is presented in **Table 1** and general structure is given in **Fig.2**

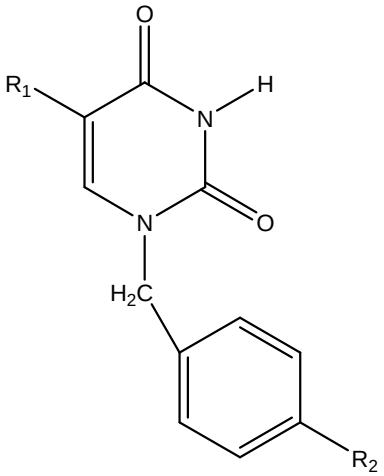


Fig.2 General Structure

Table 1:- The selected series of derivatives along with their Antimycobacterial activity data; IC_{50} (µmoles)

Molecule no.	R_1	R2	IC_{50}	PIC_{50}
1.	$-CH_3$	$-CH_2CH_2CONH_2$	89	-1.9493
2.	$-CH_3$	$-CH_2CH_2COOH$	55	-1.7403
3.	$-CH_3$	$-CH=CHCONH_2$	195	-2.2900
4.	$-CH_3$	-CH=CHCOOH	39	-1.5910
5.	$-CH_3$	$-(CH_2)_3COOH$	13	-1.1139
6.	$-CH_3$	$-(CH_2)_3CONH_2$	112	-2.0492
7.	$-CH_3$	$-C \equiv CCH_2CH_2OH$	70	-1.8450
8.	$-CH_3$	$-(CH_2)_4OH$	51	-1.7075
9.	-H	$-(CH_2)_3COOH$	202	-2.3053
10.	-Br	$-(CH_2)_3COOH$	10	-1.0000
11.	-Cl	$-(CH_2)_3COOH$	6.5	-0.8129
12.	-Br	$-(CH_2)_3CONH_2$	39	-1.5910
13.	-Cl	$-(CH_2)_3CONH_2$	39	-1.5910
14.	$-CH_3$	$-(CH_2)_4COOH$	58	-1.7634
15.	$-CH_3$	$-(CH_2)_4CONH_2$	55	-1.7403
16.	-Br	$-(CH_2)_4COOH$	38	-1.5797
17.	-Cl	$-(CH_2)_4COOH$	16	-1.2041
18.	-Br	$-(CH_2)_4CONH_2$	35	-1.5440
19.	-Br	$-(CH_2)_5COOH$	34	-1.5314
20.	-Cl	$-(CH_2)_5COOH$	23	-1.3617
21.	-Br	$-(CH_2)_5CONH_2$	19.5	-1.2900
22.	-Cl	$-(CH_2)_5CONH_2$	26	-1.4149
23.	-Br	$-(CH_2)_3COOCH_3$	52	-1.7160
24.	-Cl	$-(CH_2)_3COOCH_3$	59	-1.7708
25.	$-CH_3$	-Br	38	-1.5797

In order to correlate the free energy changes of in vivo interactions of reported compounds with the target enzyme, the IC_{50} values were converted to pIC_{50} according to equation 1[14]

$pIC_{50} = -\log IC_{50}$ ----------**Equation 1**

The G-QSAR studies were performed using kNN MFA model using Simulating Annealing and Genetic Algorithm variable selection method. For kNN MFA analysis [15], appropriate training set of molecules was used to generate QSAR models. The reliability and precision of the results of the QSAR analysis are sensitive to the variations in the structural and biological activity data. As a matter of the fact that the training set must be designed in such a way that a maximum of information can be obtained with the minimum number of compounds in order to speed up the analysis process and improve the predictiveness of the model. The sphere exclusion method was used for the selection of molecules in training and test set and dissimilarity value was accordingly set.

G-QSAR methodology was implemented in VLife MDS software [16]. Molecular fragmentation is a prerequisite to perform G-QSAR and it was done using template based method. The template used to define the dummy atoms is given in **Fig.3**

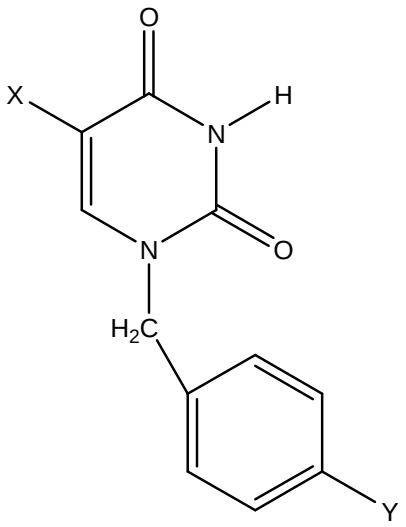


Fig.3: The template

The template search in set of molecules is performed by the software and then molecule gets fragmented where the dummy atom matches with atom of molecule.

There are two different substitution sites and for the G-QSAR of this set the molecules are divided into fragments composed of various substituents. Various 2D Descriptors were then calculated for each fragment of the molecule. Invariable columns were removed. The G-QSAR model was then build by selection of training and test set using Sphere Exclusion (SE) method [17-19] and they are mentioned in **Table 2 and Table 3 respectively**.

Table 2 :- Training set of benzyl pyrimidine derivatives along with biological activity, predicted activity data and residuals obtained thereof for SA-kNN-MFA & GA –kNN-MFA models

Sr. No.	Molecule no.	IC_{50} (µmoles)	pIC_{50}	SA-kNN-MFA		GA-kNN-MFA	
				Predicted activity	Residual activity	Predicted activity	Residual activity
1.	11	6.5	-0.8129	-1.0000	0.1871	-1.0000	0.1871
2.	13	39	-1.5910	-1.5910	0	-1.5910	0
3.	14	58	-1.7634	-1.4150	-0.3484	-1.7710	0.0076
4.	25	38	-1.5797	-1.7080	0.1288	-1.4150	-0.1647
5.	18	35	-1.5440	-1.3620	-0.182	-1.7080	0.164
6.	1	89	-1.9493	-1.5910	-0.3583	-2.2900	0.3407
7.	22	26	-1.4149	-1.5910	0.1761	-1.5800	0.1651
8.	23	52	-1.7160	-1.7710	0.055	-1.7710	0.055
9.	24	59	-1.7708	-1.7160	-0.0548	-1.7630	0.008
10.	16	38	-1.5797	-1.4750	-0.1647	-1.5800	0.0003
11.	3	195	-2.2900	-1.9490	-0.341	-1.9490	0.341
12.	4	39	-1.5910	-1.7080	0.117	-1.7080	0.117
13.	7	70	-1.8450	-1.9490	0.104	-2.2900	0.445
14.	8	51	-1.7075	-1.5910	-0.1165	-1.5910	-0.117
15.	9	202	-2.305	-2.0490	-0256	-1.7080	-0.597

Table 3:- Test set of benzyl pyrimidine derivatives along with biological activity, predicted activity data and residuals obtained thereof for SA kNN-MFA &GA kNN-MFA models

Sr. No.	Molecules No.	IC_{50} (μmoles)	PIC_{50}	SA-kNN-MFA		GA -kNN-MFA	
				Predicted activity	Residual activity	Predicted activity	Residual activity
1.	10	10	-1.0000	-1.1140	0.114	-1.2040	0.2040
2.	12	39	-1.5910	-1.5910	0	-1.5440	-0.047
3.	2	55	-1.7403	-1.7080	-0.0323	-1.5910	-0.1493
4.	21	19.5	-1.5314	-1.2900	-0.2414	-1.4150	-0.1164
5.	15	55	-1.7403	-2.0490	0.3046	-1.5910	-0.1493
6.	5	13	-1.1140	-1.000	-0.114	-1.5440	0.43
7.	6	112	-2.0490	-1.7630	-0.286	-2.2900	0.18
8.	17	16	-1.2040	-1.4150	0.211	-1.000	-0.2040

The Simulating Annealing (SA) and Genetic Algorithm variable selection methods [17] were selected and model was build using Multiple Linear Regression method (MLR).

The Statistical Results of Group QSAR equation obtained by using Simulating Annealing and Genetic Algorithm method are shown in **Table 4** and **Table 5** respectively.

Table.4:- Statistical Results of Group QSAR equation obtained by using Genetic Algorithm method:-

Sr.No.	Statistical parameters	Result	Contributing Descriptors(contribution)
1.	r^2	0.7926	R1-polarizabilityAHC R2-4pathClusterCount R1-ChlorinesCount R2-polarizabilityAHP R2-XlogP
2.	q^2	0.5876	
3.	Pred r^2	0.5120	
4.	Pred r^2SE	0.2000	
5.	F-Test	9.8797	
6.	Alpha Rand q^2	0.05	
7.	Best Rand q^2	0.42419	
8.	Z-score q^2	1.90452	

Table 5:- Statistical Results of Group QSAR equation obtained by using Simulating Annealing method:-

Sr.No.	Statistical parameters	Result	Contributing Descriptors(contribution)
1.	r^2	0.8028	R2-SsNH2E-index
2.	q^2	0.5392	
3.	Pred r^2	0.5563	R2-SssOcount
4.	Pred r^2SE	0.3296	
5.	F-Test	10.1783	R1-polarizabilityAHC
6.	Alpha Rand q^2	0.050	
7.	Best Rand q^2	0.530	R1-ChlorinesCount
8.	Z-score q^2	1.91765	

Based on the results G- QSAR Equation generated with Multiple Linear Regression Method (MLR) using Genetic Algorithm method was selected on the basis of Pred r^2SE, (error) which should be less value.

G- QSAR Equation generated using Multiple Linear Regression Method (MLR) using Genetic Algorithm method is presented in **Eq.2**

pIC_{50} = 0.2297(±0.0622) R1-polarizabilityAHC - 0.2262(±0.0844) R2-4pathClusterCount
+ 0.3669(±0.1050) R1-ChlorinesCount + 0.0081(±0.0084) R2-polarizabilityAHP
+ 0.1803(±0.0775) R2-XlogP - 2.4367 ---------- **Equation 2**

Thegraphical results are expressed using following figure 4

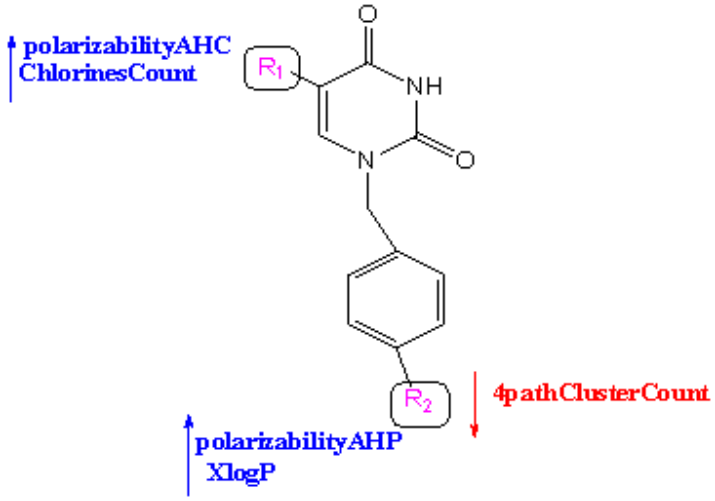


Fig.4 Graphical results of G-QSAR

Design of New Chemical Entities

Using the results of G- QSAR studies, the pharmacophore, shown in the **Fig. 5**, was designed which gives the complete details of the requirements around the pyrimidine pharmacophore.

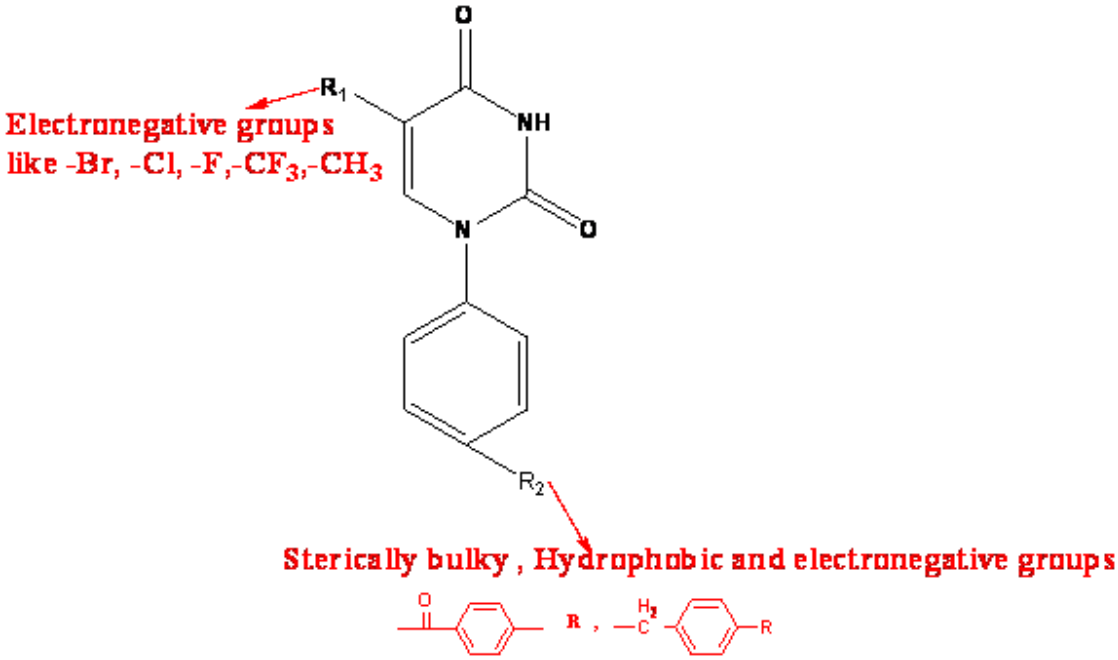


Fig. 5 Designed Pharmacophore

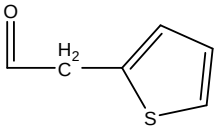
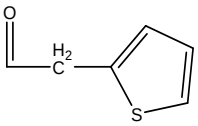
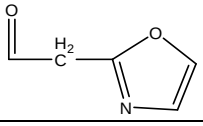
The designed pharmacophore was used for the design of potent NCEs for selective inhibition of Thymidine monophosphate kinase enzyme as shown in **Table 6**.

The NCEs were designed based on the information generated by G-QSAR studies. Following filters were used while generating Combilib, to ensure drug like pharmacokinetic profile of the designed NCEs.

A= Number of Hydrogen bond Acceptors is less than 10,
D= Number of Hydrogen bond donors is less than 5,
R= Number of Rotatable bonds is less than 10,
X= X log P not more than 5
W= Molecular weight is less than 500g/ mol,
S= Polar Surface Area is less than or equal to 140 Å

The generated Combilib was analyzed by Lipinski's screen. Compounds qualifying all required parameters set for Lipinski's screen filter are indicated by word ADRXWS in **Table 6**. Table columns containing the Lipinski' score and other column containing the strings of alphabets ADRXWS indicate that those requirements are satisfied by that corresponding compound. If molecules do not satisfy all 6 criteria, then those numbers of alphabet strings will be missing from the column and the screen score will be reduced indicating lesser pharmacokinetic compatibility for that compound. Results indicate that designed NCEs are satisfying all the parameters set for Lipinski's screen. The activity of the designed NCEs was predicted using the equation generated using the Multiple Linear Regression method.

Table.6:- Result of NCEs using CombiLib tool of VLife MDS with Predicted activity

Sr. No.	R_2	R_1	Screen Result	Screen Score	Predicted Activity
1.	-4-NH_2 Phenyl	-Cl	ADRXWS	6	-0.986
2.	-4-NO_2 Phenyl	-Cl	ADRXWS	6	-1.290
3.	-4-OCH_3 Phenyl	-Cl	ADRXWS	6	-1.327
4.	-4-$CONH_2$ Phenyl	-Cl	ADRXWS	6	-1.862
5.	-4-NH_2 Phenyl	-H	ADRXWS	6	-1.8766
6.	-4-OCH_3 Phenyl	-H	ADRXWS	6	-1.2291
7.	-4-Cl Phenyl	-H	ADRXWS	6	-1.0483
8.	-4-NO_2 Phenyl	-H	ADRXWS	6	-1.0813
9.		-Cl	ADRXWS	6	-1.170
10.	-4-OCH_3 Phenyl	-CH_3	ADRXWS	6	-2.229
11.	-4-OH Phenyl	-CH_3	ADRXWS	6	-2.307
12.		-Br	ADRXWS	6	-2.322
13.		-F	ADRXWS	6	-3.284

RESULTS AND DISCUSSION

Thus **G-** QSAR model was developed according to Multiple Linear Regression Method (MLR) using Genetic Algorithm method and Simulating Annealing method. As the former method resulted in better predictability, the contribution of descriptors in this method was considered. The R1-polarizabilityAHC and R2-polarizabilityAHP are Topoloical descriptors. Positive contribution of these descriptors viz. R1-polarizabilityAHC and R2-polarizabilityAHP indicates that polarizable substituents are preferred at those sites. The R2-4pathClusterCount is a topological descriptor based on the molecular graph of a molecule. Negative contribution of this descriptor indicates branching within the substitution contributes less for the activity. The R1-ChlorinesCount is a physicochemical descriptor of sub-class element count. This descriptor signifies number of chlorine atoms in compound and positive contribution indicates chlorine is a preferred substituent. The R2-XlogP is a Thermodynamic descriptor which signifies ratio of solute concentration in octanol and water and generally termed as Octanol water partition Coefficient. This is atom based evaluation of logP and positive contribution indicates that hydrophobicity is preferred.

Contribution(%) of each descriptor in antiTB activity studied using Genetic Algorithm method by G-QSAR model is shown in **Fig No.6** .

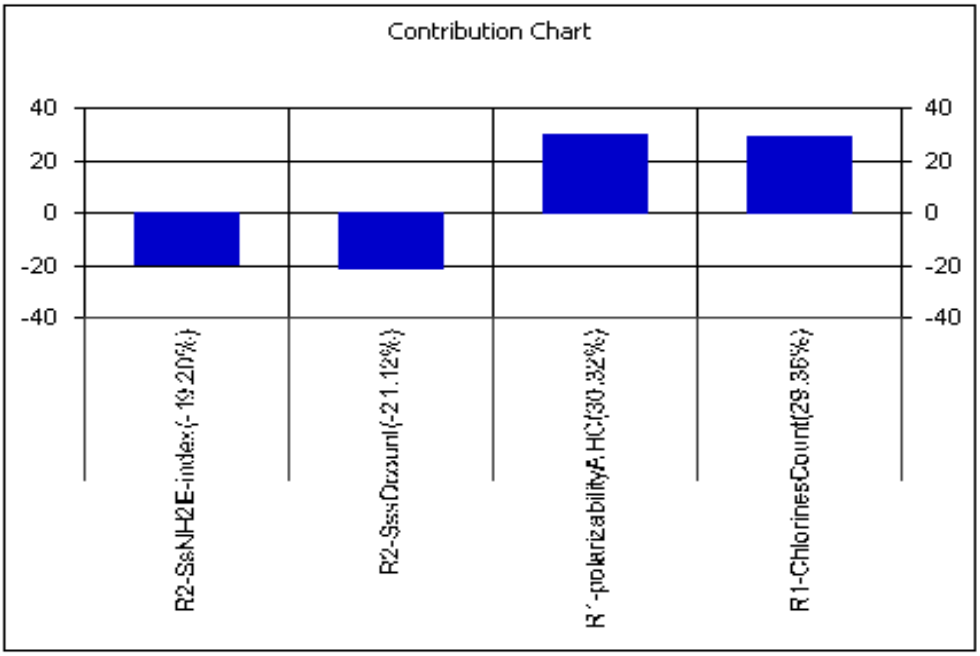


Fig.6: **Contribution (%) of each descriptor**

Thus interpretation of descriptor contribution obtained and the resulting equation using Genetic Algorithm method, was utilized to predict the activity of newly designed molecules based on the positive and considering the contribution of negatively contributing descriptors. Results indicate that designed NCEs are satisfying all the parameters set for Lipinski's screen Thus the proposed G-QSAR methodology has provided an approach for better understanding of the structure activity relationship both in terms of identifying important chemical variations at specific substitution sites and also by providing mathematical model for the prediction of

activities of the new molecules. The site specific clues along with the interpretation of descriptors provided by G-QSAR will further help us to design better molecules.

REFERENCES

[1] V Vanheusden; M lehmann; H.Froeyen; M.Dugue; Lheyerick;A; De Keukeleire;D Pochet; Sbusson;R Herdewijn. J.*Med.Chem.* , **2003**, 46, 3811-3821.
[2] V Aparna; J Jeevan; M Ravi; G Desiraju; B Gopalakrishnan. *Bioorg. Med. Chem. Lett.*, **2006**, 16, 1014-1020.
[3] K Mdluli; M Spigelman, *Current Opinion in Pharmacology*. , **2006**, 6, 459-467.
[4] MLehmann;H.Chaffotte;A Pochet; SLabesse. G. *Protein Sci*., **2001**, 10, 1195-1198.
[5] V Vanheusden; M lehmann; H.Pochet; SHerdewijin; P Van; S Calenbergh. *Bioorg.Med.Chem.Lett*., **2002**, 12, 2695-2698.
[6] A Haouz; V Vanheusden; M lehmann; H., H.Froeyen; P Herdewijin; S Calenbergh.. *tuberculosis* thymidylate kinase ,New insights into the phosphoryl transfer mechanism. *J. Biol. Chem.* ,**2003**, 278, 4963-4971.
[7] H Saito; H Tomioka. J. *Gen. Microbiol*. ,**1984**, 130, 1863.
[8] ST Cole; RJ Brosch; P Arkhill; T Garnier; C Churcher; D Harris. *Nature*,**1998**, 393, 537.
[9]D Pochet ; SDugue ; LLabesse ; G.Delepierre ; M. Lehmann.*ChemBioChem*. ,**2003**, 4, 742-747.
[10] D Douguet;M Lemann;H Labesse; G Pochet. *J.Med. Chem.* ,**2005**, 48, 2457-2468.
[11] C Gasse ; D Douguet ; V Huteau ; G Marchal ; MLehmann ;H. Pochet . *Bioorg.Med.Chem*., **2008**, 16, 6075- 6085.
[12] S Ajmani;K Jadhav; S Kulkarni. *QSAR Comb.Sci*. **2009**, 1, 36-51
[13] V. Raparti, T. Chitre, K. Bothara, and V. Kumar, *Eur. J. Med. Chem.*, 2009, **44,** 3954-3960
[14] C Hansch; T Fujita. *The Journal of the American Chemical Society*; **1964**, 86, 1616–1626.
[15] S Ajmani;K Jadhav; S Kulkarni..*Chem.Inf.Model*. ,**2006**, 46, 24- 31.
[16] VLifeMDS; Molecular Design Suite version 3.0, V-life Sciences Technologies Pvt. Ltd., Pune, India, 2004. (www.vlifesciences.com)
[17] A Jain; N Manocha; V Ravichandran; V Mourya; R.Agrawal. *Digest Journal of Nanomaterial and Biostructure*. Vol.3, No. 3, **2008**, 147-158.
[18] T Halgren. *J. Comput.Chem.* , **1996**, 17, 520-552.
[19] C Hansch; A Leo. *American Chemical Society*, **1995**, 19.