Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2011, 3 (6):53-61
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X
CODEN (USA): PCHHAX

Development of QSAR model for indoyl aryl sulfone derivatives as reverse transcriptase inhibitors

Laxman M Prajapati[1*], Vijay K Parmar[2], Manish J Patel[3], Jimish R Patel[1]

[1]Department of Pharmaceutical Chemistry, Shri B M Shah College of Pharmaceutical Education and Research, College Campus, Modasa, Gujarat, India
[2]Department of Pharmaceutical Sciences, Sardar Patel University, Vallabh Vidyanagar, Gujarat, India
[3]Department of Pharmaceutical Chemistry, S.K. Patel College of Pharmaceutical Education & Research, Ganpat University, Ganpat Vidyanagar, Kherva, Mehsana, Gujarat, India

---

ABSTRACT

QSAR model development of 39 indoyl aryl sulfones was carried out to predict reverse transcriptase inhibition activity. EC_{50} for reverse transcriptase binding was taken as biological activity. Physicochemical parameters were calculated using PaDEL descriptor software, version 2.1. Stepwise multiple linear regression analysis was applied to derive QSAR models, which were further evaluated for statistical significance and predictive power by internal and external validation. The best quantitative structure activity relationship model was selected having a correlation coefficient (R^2) of 0.835, cross-validated correlation coefficient (Q^2) of 0780 and, R^2_{pred} of 0.830. The predictive ability of the selected model was also confirmed by leave one-out cross-validation. The QSAR model indicates that the descriptors (nHBint, SaaNH, MDEO-11 and minaaaC) play an important role in enzyme binding. The information derived from the present study may be useful in the design of more potent substituted indoyl aryl sulfones.

Key words: QSAR, indoyl aryl sulfones, Non nucleoside reverse transcriptase inhibitors, Multiple linear regression, HIV.

INTRODUCTION

Human immune deficiency virus (HIV) is the causative agent for acquired immune deficiency syndrome (AIDS) which cause loss of helper T lymphocytes and heavy damage to lymphatic tissues [1]. HIV drugs mainly can be classified into three classes nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs) [2]. NNTRIs are characterized by different unrelated chemical structures. The

marketed drugs nevirapine, delaviridine, and efavirenz are significant example of such inhibitors. At the molecular level NNTRIs act by a specific allosteric effect arising from noncompetitive binding to a hydrophobic pocket, the non- nucleoside binding side (NNBS), located near the catalytic side [3,4]. Continuous efforts in this field are documented by the wide number if NNTRI described in the literature, some of which are under clinical trials [5-10]. Among them L-737,126, a benzenesulfonylindolcarboxamide endowed with potent antiviral activity and high selectivity,has been developed by Meck A. G. During the extensive structure activity relationship (SAR) studies on diaryl sulfones, first identified pyrryl sulfones as highly potent NNRTI and then the research extended to novel indoyl aryl sulfones. In particular, indole derivatives having 2-methylphenylsulphonyl or 3-methylphenylsulphonyl moieties were found to inhibit HIV-1 at nanomolar concentrations [11-13]. Further the introduction of a 3,5-dimethylphenylsulfonyl moiety led to compounds displaying high activity and selectivity. Pursuing these research consequences we have undertaken QSAR study on previously reported IAS. The aim of the study was to identify the molecular properties which increase interaction between non-nucleoside reverse transcriptase and designed compounds.

MATERIALS AND METHODS

A total of 39 indoyl aryl sulfones reported as non-nucleoside reverse transcriptase inhibitors [14, 15] were used as the data set in QSAR analysis **(Table 1)**. These molecules have found to be active against HIV-1 at nano molar concentration. The EC_{50} (µM) were converted in to molar values which again converted to negative logarithmic values to get pEC_{50} for QSAR study.

Molecules were divided into the training set (29 molecules) and test set (10 molecules) by random selection. The structures were drawn and transformed to 3D on software ChemOffice 2004 [16]. The energy minimization was performed using molecular mechanics-2 (MM2) until the root mean square (RMS) gradient value became smaller than 0.100 kcal/mol Å and then molecules were subjected to re-optimization via MOPAC (Molecular Orbital Package) method until the RMS gradient attained a value smaller than 0.0001kcal/mol Å.

Various descriptors like electronic, steric, and thermodynamic were calculated on the PaDEL descriptor software, version 2.1 **(Table 2)** [17]. Stepwise multiple linear regression method was applied for generation of QSAR model using VALSTAT program [18]. For the validation of QSAR models “Leave-one-out (LOO)” method was used, the best model was selected on the basis of various statistical parameters such as correlation coefficient (R), square of correlation coefficient (R^2), sequential Fischer test (F). The quality of the each model was estimated from the cross-validated squared correlation coefficient (Q^2), standard deviation of prediction (S_{PRESS}), Standard deviation of error of prediction (S_{DEP}). Boot-strapping square correlation coefficient (R^2_{bt}) was calculated to confirm the robustness and applicability of QSAR equation. The derived QSAR models were used for the prediction of the activity compounds in the test set, R^2_{pred} was calculated.

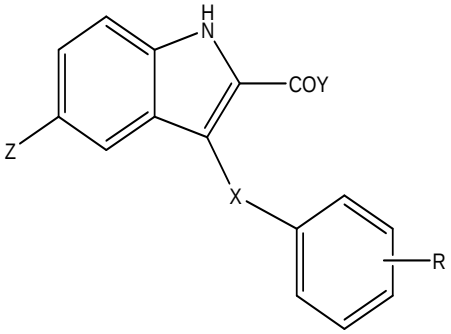

Table 1 Chemical and biological data of indoyl aryl sulfones

Compound no.	X	Y	Z	R	pEC_{50}
1	S	OEt	Cl	2-NH_2	5.638
2	S	OEt	Cl	2-NH_2-5-Cl	5.602
3	SO_2	OEt	H	H	5.432
4	SO_2	OEt	H	2-NH_2-5-Cl	5.602
5	S	NH_2	H	H	5.854
6	S	NH_2	H	2-NH_2-5-Cl	5.046
7	SO_2	OEt	Cl	2-NH_2-5-Cl	5.721
8	S	NH_2	Cl	4-CH_3	6.523
9	S	NH_2	Cl	4-F	5.854
10	S	NH_2	Cl	4-iso-Pr	5.721
11	S	NH_2	Cl	4-tert-Bu	5.097
12	S	NH_2	Cl	2,6-Cl_2	5.921
13	S	NH_2	Cl	2-NH_2-5-Cl	5.796
14	SO_2	NH_2	H	2-NH2-5-Cl	6.523
15	SO_2	NH_2	Cl	2-Me	9.0
16	SO_2	NH_2	Cl	3-Me	9.0
17	SO_2	NH_2	Cl	4-Me	8.523
18	SO_2	NH_2	Cl	4-F	7.854
19	SO_2	NH_2	Cl	4-Cl	7.959
20	SO_2	NH_2	Cl	4-iso-Pr	7.095
21	SO_2	NH_2	Cl	4-tert-Bu	6.886
22	SO_2	NH_2	Cl	2,4-Me_2	8.398
23	SO_2	NH_2	Cl	2,4-Me_2	8.398
24	SO_2	NH_2	Cl	H	9.0
25	SO_2	NH_2	Cl	2-NH_2-5-Cl	7.398
26	SO_2	NH_2	Br	3,5-Me_2	8.699
27	S	$NHNH_2$	Cl	H	6.260
28	SO_2	NH_2	COMe	3,5-Me_2	7.824
29	S	$NHNH_2$	Cl	4-Me	5.823
30	SO_2	NH_2	CO(OH)Me	3,5-Me_2	7.602
31	S	$NHNH_2$	Cl	4-F	5.301
32	S	$NHNH_2$	Cl	4-Cl	5.0
33	SO_2	$NHNH_2$	Cl	H	8.0
34	SO_2	$NHNH_2$	Cl	4-Me	8.0
35	SO_2	$NHNH_2$	Cl	4-F	7.301
36	SO_2	$NHNH_2$	Cl	4-Cl	6.495
37	SO_2	$NHNH_2$	Cl	3,5-Me_2	6.721
38	S	OEt	H	H	5.854
39	SO_2	$NHNH_2$	Cl	2-NH_2-5-Cl	6.523

Table 2 Calculated values of various descriptors for the set of compounds

Compound No.	nHBint[a]	maxaaaC[b]	MDEO-11[c]	SaaaC[d]	SaaNH[e]	minaaaC[f]
1	0	1.174	0	2.187	3.272	1.013
2	0	1.190	0	2.214	3.297	1.025
3	2	0.573	0.814	1.059	2.887	0.486
4	2	0.585	0.814	1.086	2.911	0.502
5	1	1.150	0	2.150	3.163	1.0
6	1	1.115	0	2.088	3.166	0.973
7	2	0.618	0.814	1.177	2.957	0.559
8	1	1.212	0	2.247	3.227	1.036
9	1	1.115	0	2.076	3.145	0.960
10	1	1.211	0	2.247	3.238	1.036
11	1	1.208	0	2.242	3.244	1.034
12	1	1.252	0	2.317	3.274	1.065
13	1	1.227	0	2.275	3.252	1.048
14	2	0.567	0.814	1.051	2.826	0.484
15	0	0.630	0.814	1.210	2.888	0.581
16	0	0.628	0.814	1.208	2.883	0.579
17	0	0.628	0.814	1.206	2.880	0.578
18	0	0.554	0.814	1.039	2.806	0.485
19	0	0.636	0.814	1.225	2.888	0.589
20	0	0.630	0.814	1.210	2.898	0.581
21	0	0.627	0.814	1.205	2.904	0.578
22	0	0.631	0.814	1.213	2.900	0.582
23	0	0.630	0.814	1.212	2.897	0.582
24	0	0.627	0.814	1.203	2.868	0.577
25	2	0.600	0.814	1.142	2.872	0.542
26	0	0.641	0.814	1.241	2.905	0.599
27	1	1.203	0	2.231	3.241	1.029
28	0	0.460	1.156	0.795	2.807	0.335
29	1	1.204	0	2.234	3.253	1.030
30	0	0.501	1.156	0.897	2.837	0.396
31	1	1.110	0	2.066	3.179	0.956
32	1	1.214	0	2.252	3.261	1.038
33	0	0.622	0.814	1.194	2.902	0.572
34	0	0.623	0.814	1.197	2.913	0.574
35	0	0.549	0.814	1.029	2.839	0.480
36	0	0.625	0.814	1.202	2.931	0.577
37	0	0.625	0.814	1.202	2.931	0.577
38	0	1.168	0	2.186	3.248	1.018
39	2	0.596	0.814	1.132	2.905	0.536

[a]Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 4,[b]Maximum atom-type E-State:C, [c]Molecular distance edge between all primary oxygens, [d]Sum of atom-type E-State:C, [e]Sum of atom-type E-State:NH, [f]Minimum atom-type E-State:C

The Z-score was calculated for the detection of outliers. Z-score can be defined as absolute difference between the value of the model and the activity field, divided by the square root of the mean square error of the data set. Any compound which shows a value of Z-score higher than 2.5, during generation of a particular QSAR model, was considered as outlier. Finally, the derived QSAR models were used for the prediction of the activity of the compounds in the test set and the external validation parameter, predictive R^2 (R^2_{pred}) was calculated for evaluating the predictive capacity of the model.

RESULT AND DISCUSSION

In present study authors tried to develop QSAR model to establish the correlation between physicochemical parameters and reverse transcriptase inhibiting capacity. A reported data set of 39 indoyl aryl sulfones derivatives was used in present study.

When data set was subjected to sequential multiple linear regression analysis several equations were obtained. Out of these three most statistically significant equations were considered as significant. The statistical parameters for these models are shown in **Table 3.**

BA= [35.8094(± 16.6583)] + nHBint4 [-0.523857(± 0.296371)] + MDEO-11 [7.44356(± 4.1953)] + SaaaC [9.58861(± 4.35596)] + SaaNH [-15.7002(± 6.66672)]

BA= [30.1714(± 17.3404)] + nHBint4 [-0.525761(± 0.326358)] + MDEO-11 [5.48921(± 3.9569)] + SaaNH [-12.2278(± 6.32326)] + minaaaC [15.1991(± 8.1875)]

BA= [37.5564(± 18.5803)] + nHBint4 [-0.647995(± 0.299808)] + maxaaaC [15.8681(± 8.74908)] + MDEO-11 [6.34406(± 4.43381)] + SaaNH [-15.486(± 7.54577)]

Table 3 QSAR statistics of significant equations[#]

Parameters	Model no.1	Model no.2	Model no.3
N Train	29	29	29
N test	10	10	10
NV	4	4	4
R	0.926	0.914	0.912
R^2	0.857	0.835	0.832
Variance	0.246	0.284	0.289
Std	0.496	0.533	0.538
F	35.863	30.258	29.630
R^2_{bt}	0.854	0.839	0.842
Chance	<0.001	<0.001	<0.001
Q^2	0.812	0.780	0.766
S_{PRESS}	0.569	0.614	0.635
S_{DEP}	0.517	0.559	0.577
R^2_{pred}	0.774	0.830	0.656

#N Train= number of training set, N Test= number of test set, NV= number of variables, R= coefficient of correlation, R^2= squared correlation coefficient, Std= standard deviation of estimation, F= Fischer's value, R^2_{bt}= boot-strapping square correlation coefficient, Q^2=cross-validated squared correlation coefficient, SPRESS= predictive residual sum of square, S_{DEP} = standard error of prediction. R^2 = predicted coefficient of correlation

All the three model have good correlation and internal predictivity, model no.2 was considered best model because of its predictivity on external compounds. The intercorrelation of some descriptor was found high (**Table 4)**, which is could be due synergistic interaction of descriptors. Furthermore, the multicollinearity that results from employing correlated descriptors is not as serious a problem as is often assumed. There are several examples of cases in which pairs of highly correlated, poorly performing single-parameter descriptors produce significant regression equations [19].

Table 4 Correlation matrix for the inter-correlation of structural descriptors and their correlation with the activity

Parameters	pEC_{50}	nHBint4	maxaaaC	MDEO-11	SaaaC	SaaNH	minaaaC
pEC_{50}	1						
nHBint4	0.579	1					
maxaaaC	0.680	0.080	1				
MDEO-11	0.708	0.126	0.992	1			
SaaaC	0.667	0.066	0.999	0.992	1		
SaaNH	0.719	0.090	0.986	0.974	0.985	1	
minaaaC	0.651	0.049	0.996	0.989	0.999	0.981	1

Table 5 Activity of training set for Model No. 2

Compound no.	Observed activity	Calculated activity	Predicted activity
1	5.638	5.560	5.537
2	5.602	5.436	5.370
3	5.432	5.680	5.753
4	5.602	5.617	5.621
5	5.854	6.170	6.243
6	5.046	5.724	5.855
7	5.721	5.930	6.021
8	6.523	5.926	5.837
9	5.854	5.784	5.763
13	5.796	5.802	5.803
14	6.523	6.387	6.338
16	9.0	8.195	8.092
17	8.523	8.221	8.181
19	7.959	8.278	8.326
20	7.097	8.031	8.129
21	6.886	7.915	8.011
23	8.398	8.053	8.016
24	9.0	8.336	8.227
27	6.260	5.641	5.567
28	7.824	7.283	7.017
29	5.824	5.518	5.478
31	5.301	5.298	5.297
32	5.0	5.542	5.627
33	8.0	7.850	7.837
34	8.0	7.735	7.712
35	7.301	7.216	7.170
36	6.495	7.567	7.682
38	5.854	5.925	5.938
39	6.523	6.219	6.138

Table 6 Activity of test set for Model No. 2.

Compound no.	Observed activity	Predicted activity
10	5.805	5.721
11	5.696	5.721
12	5.805	5.921
15	8.160	9.0
18	7.702	7.854
22	8.045	8.398
25	6.700	7.398
26	8.225	8.699
30	7.845	7.602
37	7.568	6.721

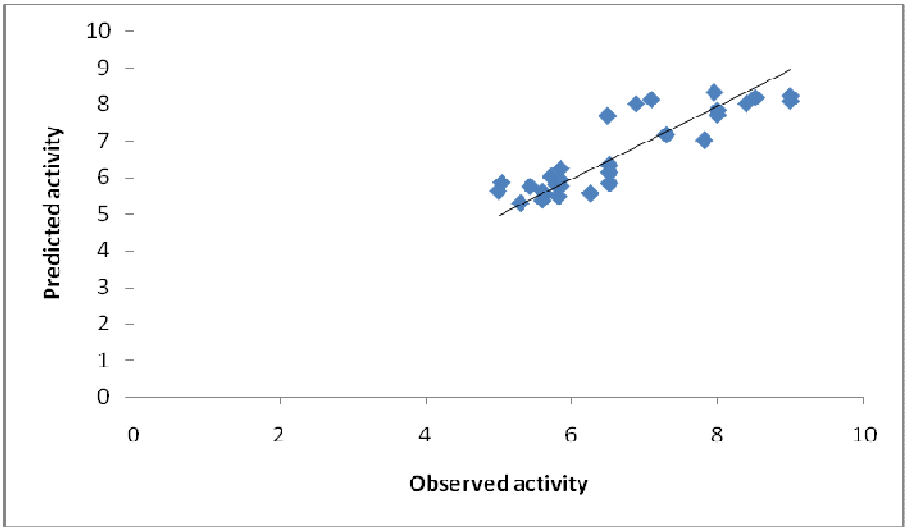

Fig. 1 Graphs of actual versus predicted activity (pEC_{50}) of the training sets for the model 2.

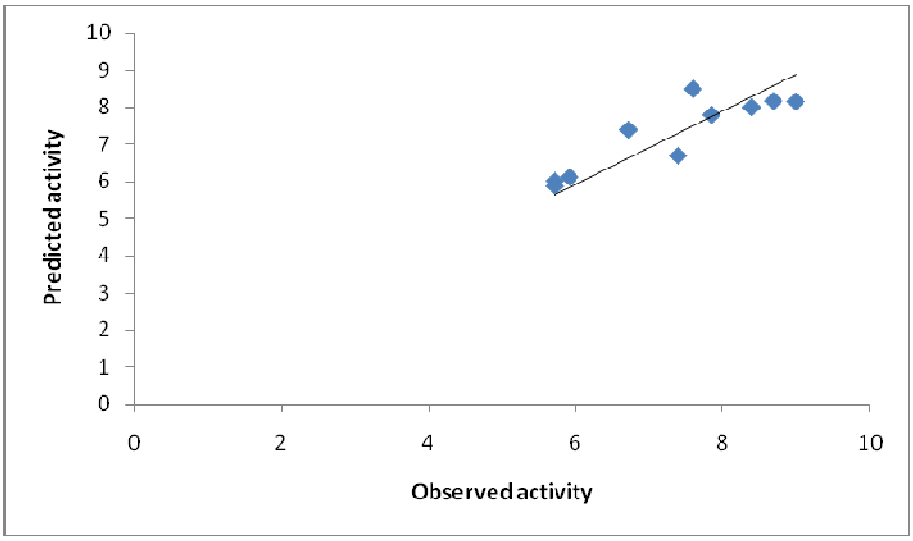

Fig. 2 Graphs of actual versus predicted activity (pEC_{50}) of the test for the model 2.

Model 2 shows a good correlation coefficient (R) of 0.914 between the descriptors nHBint4, MDEO-11, SaaNH, minaaaC and pEC_{50} for reverse transcriptase. The R^2 of 0.835 explains 83.5 % of the variance in biological activity. This model also shows significance level more than 95% against tabulated value F=30.2580, with a low standard deviation of estimation 0.533, manifest of accuracy of the model. The stability of model-2 judged by leave-one-out procedure is good (Q^2= 0.780) suggesting that the models will be useful for meaningful predictions. The robustness of model was shown by magnitude of the R^2_{bt} (0.839), which was near to conventional R^2 (0.835). Further support in this regard is obtained from the low values of the cross-validation parameters S_{PRESS} and S_{DEP}. The predicted R^2 value of the test set was 0.830, indicating excellent predictive ability of model 3. The observed, calculated and predicted values of fold selectivity are shown in **Table 5 and Table 6**. The correlation between observed and predicted activity (LOO) of training set is shown in **Fig 1**. The correlation between observed and predicted activity of training test set is shown in **Fig 2**.

Negative contribution of nHBint and SaaNH in biological activity indicates increased value of this parameter increases the pEC_{50} value making compound more potent against HIV-1. Positive contribution of MDEO-11 and minaaaC in biological activity indicates increased value this parameter decreases pEC_{50} and thus decreases potency against HIV-1.

CONCLUSION

It can be concluded, by decreasing or increasing these parameter values binding affinity of indoyl aryl sulfones to reverse transcriptase can be increased. The equation will help to develop new compounds in indoyl aryl sulfones series with high potency.

REFERENCES

[1] WHO/UNAIDS AIDS Epidemic update.December **2003**.
[2] I.G. Williams, *Int. J. Clin. Pract.*, **2003**, 57, 890-897.
[3] C. Tantillo, J. Ding, A. Jacobo-Molina, R. G. Nanni, P. L. Boyer, S. H. Hughes, R. Pauwels, K. Andries, P. A. Janssen, E. Arnold, *J. Mol. Bio.*,**1994**, 243, 369-387.
[4] R. A. Spence, W. M. Kati, K. S. Anderson, K. A. Johnson, *Science*, **1995**, *267*, 988-993.
[5] S. D. Young, *Annu. Rep. Med.Chem.*, **2003**, 38, 173-182.
[6] R. W. Buckheit, *Exp. Opin. Invest. Drugs*, **2001**, 10, 1423-1442.
[7] M. Artico, *Farmaco.*, **1996**, 51, 305-331.
[8] D. L. Flynn, *Abstracts of 226th ACS National Meeting*, **2003**, 127-130.
[9] E. De Clercq, *Med. Res. Rev.*, **2002**, 22, 531-565.
[10] E. De Clercq, *Chem. Biodiversity*, **2004**, 1, 44-64.
[11] M. Artico, R. Silvestri, G. Strefancich, S. Masaa, E. Pagnozzi, D. Musu, F. scintu, E. Pinna, E. Tinti, P. La Colla, *Arch pharm.*, **1995**, 328, 223-229.
[12] M. Artico, R. Silvestri, S. Masaa, A. G. Loi, S. Corrias, G. Piras, La Colla, *J. Med .chem.*, **1996.** 39, 522-530.
[13] M. Artico, R. Silvestri, , E. Pagnozzi, B. Bruno, B. Novellino, G. Greco, S. Masaa, A. Ettorre, A. G. Loi, , F. scintu, P. La Colla, *J. Med .chem.*, **2000,** 43, 1886-1891.
[14] S. D. Young, M. C. Amblard, S. F. Brichter, V. E. Grey, L. O. Tran, W. C. Lumma, J. R. Huff, W. A. Schleif, E. E. mini, J. A. O`Brien, D. J. Pettibone, *Bio.org. Med.Chem.Lett.*, **1995**, 5, 491-496.

[15] R. Silvestri, G. De Martino, G. La Regina, M. Artico, S. Massa, L, Vargiu, M. Mura, A. G. Loi, T. Marceddu, P. La Colla, *J. Med. Chem.*, **2003**, 46, 2482-2493.
[16] CS Chem Office, Version 8.0, Cambridge Soft Corporation, Software Publishers Association, 1730 M Street, NW, Suite 700, Washington, D.C. 20036.
[17] CW.Yap, PaDEL-Descriptor, *Journal of Computational Chemistry*, **2010**, 32, 1466-1474.
[18] AK. Gupta, BM. Arockia, SG. Kaskhedikar, *Ind. J. Pharm. Sci*., **2004**, 66, 396-402.
[19] C. Stephen, O. Peterangel, G.Paul, D. Seybol, *International Journal of Quantum Chemistry*, **2010**, 96, 1-9.