

QSAR study of novel 1,1,3-trioxo[1,2,4]-thiadiazine (TTDs) analogues as potent anti-HIV agents

V. Ravichandran,^a V. K. Mourya,^b and R. K. Agrawal^{a*}

^aPharmaceutical Chemistry Research Laboratory, Department of Pharmaceutical Sciences,
Dr. H. S. Gour Vishwavidyalaya, Sagar (MP) – 470 003, India

^bGovernment College of Pharmacy, Aurangabad (MH), India

E-mail: dragrawal2001@yahoo.co.in

Abstract

In pursuit of better anti-HIV agents, QSAR studies were performed on a series of 1, 1, 3 - trioxo [1, 2, 4] - thiadiazine (TTDs) analogues. Stepwise multiple linear regression analysis was performed to derive QSAR models which were further evaluated for statistical significance and predictive power by internal and external validation. The best QSAR model was selected, having correlation coefficient (r) = 0.864, standard error of estimation (SEE) = 0.443 and cross validated squared correlation coefficient (q^2) = 0.623. The predictive ability of the selected model was also confirmed by leave 20 % out cross validation. The QSAR model indicates that the thermodynamic descriptors (molar refractivity), electronic descriptor (ionization potential), Hansen-polarity and hydrogen bond donor play an important role for the anti-HIV activities. The results of the present study may be useful on the designing of more potent TTDs analogues as anti-HIV agents.

Keywords: QSAR, 1,1,3-trioxo[1, 2, 4]thiadiazine (TTDs) analogues, anti-HIV

Introduction

HIV- 1 (Human Immunodeficiency Virus Type-1) is the pathogenic retrovirus and causative agent of AIDS or AIDS- related complex (ARC).^{1,2} When viral RNA is translated into a polypeptide sequence, it is assembled in a long polypeptide chain, which includes several individual proteins namely, reverse transcriptase, protease, integrase, etc. Before these enzymes become functional, they must be cut from the longer polypeptide chain. HIV is a retrovirus of the lentivirus family.³ The three viral enzymes of HIV play an important role in the virus replication cycle. Among them, viral reverse transcriptase (RT) catalyzes the formation of proviral DNA from viral RNA, the key stage in viral replication. Its central role in viral replication makes RT a prime target for anti-HIV-therapy.⁴

Two main categories of HIV RT inhibitors have been discovered to date. The first category of inhibitors is nucleoside analogues (e.g., AZT, 3TC, ddI, ddC) and the second category of inhibitors is nonnucleoside analogues. Nevirapine, delaviridine and efavirenz are the only nonnucleoside reverse transcriptase inhibitors (NNRTI) that have received regulatory approval. Several NNRTIs (MKC442, Troviridine, S-1153/ AG1549, PNU142721, ACT and HBY1293/GW420867X) are currently undergoing clinical trials. Efavirenz was the first potent anti-HIV drug to be approved by FDA and studies have shown that efavirenz penetrates into the cerebrospinal fluid, a common viral sanctuary. The therapeutic efficacy of the drug is mainly restricted due to the development of viral resistance associated with mutations that includes K103N, L100I and Y188L.⁵ Comparative molecular field analysis (CoMFA) and Comparative molecular similarity indices analysis (CoMSIA) are powerful and versatile tools to build and design an activity model (QSAR) for a given set of molecules in rational drug design and related applications,^{6,7} HIV-1 reverse transcriptase inhibitors,⁸⁻¹⁰ HIV-1 protease inhibitors¹¹⁻¹⁵ and HIV-1 integrase inhibitors¹⁶⁻¹⁸ and gp 120 envelope glycoprotein.¹⁹

In search of effective TTDs analogues with minimal viral resistance problems, Arranz M. E. et al.,²⁰ synthesized and evaluated a novel set of TTDs for their anti-HIV activity. Standard deviation of anti-viral activity and the structural difference between the compounds in the selected series are high. It insists as to select these series of compounds for our QSAR studies. All the anti-HIV activities used in the present study were expressed as $pEC_{50} = -\log_{10} EC_{50}$, where EC_{50} is the micro molar concentration of the compounds producing 50% reduction in the cytopathic effect caused by the virus is stated as the means of at least two experiments. EC_{50} values were assessed by MTT assay. The compounds which were not showing confirmed anti-HIV activity in the above cited literature have not been taken for our study.

Results and Discussion

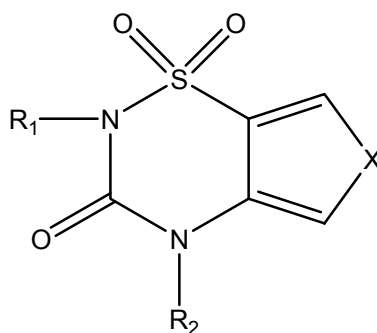
A data set of 33 compounds for anti-HIV activity was used for the present QSAR study. The QSAR studies of the TTDs series resulted in several QSAR equations. The two best equations are

$$pEC_{50} = -5.789 (\pm 2.291) + 0.174 (\pm 0.145) HP + 2.402 (\pm 1.009) HD + 0.037 (\pm 0.016) MR + 0.691 (\pm 0.286) IH \quad (1)$$

$n = 33$, $r = 0.734$, $r^2 = 0.539$, $SEE = 0.595$, $F(4, 28) = 8.19$, $P < 0.001$, $q^2 = 0.399$, $SDEP = 0.641$, $S_{PRESS} = 0.725$, $PRESS = 13.15$.

$$pEC_{50} = 13.029 (\pm 9.966) + 0.282 (\pm 0.149) HP + 3.893 (\pm 1.234) HD + 0.039 (\pm 0.015) MR - 2.139 (\pm 1.105) IP + 0.806 (\pm 0.279) IH \quad (2)$$

$n = 33$, $r = 0.772$, $r^2 = 0.596$, $SEE = 0.568$, $F(5, 27) = 7.95$, $P < 0.001$, $q^2 = 0.435$, $SDEP = 0.624$, $S_{PRESS} = 0.720$, $PRESS = 12.48$.

Table 1. Structures, physicochemical parameters and indicator variable of TTDs derivatives

Compd. No.	R ₁	R ₂	X	HP	HD	MR	IP	IH
1	<i>meta</i> -Cl-Ph	Et	S	7.635	0.038	91.558	9.318	0
2	Bn	Me	S	7.383	0.038	86.691	9.386	0
3	Bn	Et	S	6.862	0.037	91.309	9.186	0
4	Bn	n-Pr	S	6.41	0.038	95.927	9.201	0
5	Bn	Bn	S	5.901	0.065	110.798	9.24	0
6	Bn	CH ₂ CN	S	8.698	0.33	93.525	9.629	0
7	Bn	CH ₂ CH=CH ₂	S	6.62	0.055	95.46	9.193	0
8	Bn	CH ₂ CH≡CH	S	6.85	0.44	93.925	9.524	0
9	Bn	CH ₂ CH ₂ CN	S	8.102	0.037	98.143	9.48	0
10	3-picolyl	CH ₂ CN	S	9.785	0.33	92.847	9.533	0
11	Bn	<i>m</i> -Cl-Ph	S	6.517	0.065	111.047	9.224	0
12	Bn	<i>p</i> -Cl-Ph	S	6.517	0.065	111.047	9.202	0
13	Bn	CH ₂ COOEt	S	6.253	0.038	103.932	9.389	0
14	<i>ortho</i> -Cl- Bn	Et	S	7.107	0.034	96.176	9.203	0
15	<i>meta</i> -Cl- Bn	Me	S	7.383	0.038	86.691	9.257	1
16	<i>meta</i> -Cl- Bn	Et	S	7.107	0.036	96.176	9.289	1
17	<i>meta</i> -Cl- Bn	CH ₂ CN	S	8.844	0.33	98.392	9.316	0
18	<i>meta</i> -Cl- Bn	CH ₂ CH≡CH	S	7.095	0.43	98.792	9.542	1
19	<i>meta</i> -Br-Bn	CH ₂ CN	S	8.69	0.33	101.29	9.668	1
20	<i>meta</i> -F-Bn	CH ₂ CN	S	8.981	0.33	93.375	9.665	1
21	3,5-di fluoro-Bn	CH ₂ CN	S	9.26	0.32	93.225	9.712	1
22	Phenethyl	Et	S	6.41	0.036	95.927	9.218	0
23	Phenethyl	CH ₂ CN	S	8.102	0.33	98.143	9.534	0
24	Phenethyl	Bn	S	5.563	0.064	115.416	9.241	0
25	2-picolyl	Et	S	7.96	0.029	90.631	9.168	0
26	2-picolyl	CH ₂ CN	S	9.785	0.32	92.847	9.572	0
27	2-picolyl	Bn	S	6.804	0.057	110.12	9.21	0
28	Bn	<i>o</i> -Cl-Ph	S	6.517	0.063	111.047	9.237	0

29	Bn	Et	S	6.862	0.036	91.309	9.374	0
30	Bn	CH ₂ CN	S	8.698	0.33	93.525	9.686	0
31	Bn	Bn	S	5.901	0.063	110.798	9.38	0
32	Bn	Bn	C-Cl	8.844	0.33	98.392	9.653	0
33	Bn	CH ₂ CN	C-Cl	6.129	0.06	115.665	9.284	0

HP – Hansen polarity, HD – Hydrogen bond donor, MR – Molar refractivity, IP – Ionization potential and IH – Indicator variable (having value 1 if *meta* halogenated benzyl ring is present at R₁ position).

Both the models have three outliers (compounds 2, 28 and 33) because their residual values exceeded twice the standard error of estimate. When these outliers have been removed from the data set, we have got two highly significant equations 3 and 4.

$$\text{pEC}_{50} = -6.779 (\pm 2.067) + 0.181 (\pm 0.121) \text{HP} + 2.602 (\pm 0.844) \text{HD} + 0.046 (\pm 0.015) \text{MR} + 0.744 (\pm 0.238) \text{IH} \quad (3)$$

$n = 30$, $r = 0.821$, $r^2 = 0.673$, $\text{SEE} = 0.493$, $F(4, 25) = 12.91$, $P < 0.001$, $q^2 = 0.558$, $\text{SDEP} = 0.536$, $\text{S}_{\text{PRESS}} = 0.578$, $\text{PRESS} = 8.35$.

$$\text{pEC}_{50} = 13.029 (\pm 9.966) + 0.282 (\pm 0.149) \text{HP} + 3.893 (\pm 1.234) \text{HD} + 0.039 (\pm 0.015) \text{MR} - 2.139 (\pm 1.105) \text{IP} + 0.806 (\pm 0.279) \text{IH} \quad (4)$$

$n = 30$, $r = 0.864$, $r^2 = 0.747$, $\text{SEE} = 0.443$, $F(5, 24) = 14.14$, $P < 0.001$, $q^2 = 0.623$, $\text{SDEP} = 0.495$, $\text{S}_{\text{PRESS}} = 0.545$, $\text{PRESS} = 7.13$.

In the above equations n is the number of compounds used to derive the model and values in parentheses are the 95% confidence limit of respective coefficient. We extended our study for five-parametric correlations as they are permitted for a data set of 33 compounds in accordance with the lower limit of rule of thumb. The calculated and predicted (LOO) activities of the compounds by the above models are shown in Table 2.

Model-3 shows good correlation coefficient (r) of 0.821 between descriptors (HP, HD, MR and IH) and anti-HIV activity. Squared correlation coefficient (r^2) of 0.673 explains 67.3% variance in biological activity. This model also indicates statistical significance > 99.9% with F values $F(4, 25) = 12.91$. Cross validated squared correlation coefficient of this model was 0.558, which shows the good internal prediction power of this model.

Model-4 shows good correlation coefficient (r) of 0.864 between descriptors (HP, HD, MR, IP and IH) and anti-HIV activity. Squared correlation coefficient (r^2) of 0.747 explains 74.7% variance in biological activity. This model also indicates statistical significance > 99.9% with F values $F(5, 24) = 14.14$. Cross validated squared correlation coefficient of this model was 0.623, which shows the good internal prediction power of this model.

Table 2. Observed, calculated and predicted (LOO) activity of TTDs derivatives

Compd. No.	Obs. Act. ²⁰ (pEC ₅₀) ^a	Model-3		Model-4	
		Cal. Act.	Pred. Act.	Cal. Act.	Pred. Act.
1	-1.787	-1.082	-0.992	-1.164	-1.081
2	-0.431	*	*	*	*
3	-0.38	-1.236	-1.365	-1.108	-1.229
4	-0.785	-1.102	-1.138	-1.042	-1.072
5	-0.342	-0.439	-0.454	-0.417	-0.429
6	0.06	-0.039	-0.051	-0.223	-0.266
7	-0.799	-1.042	-1.065	-0.912	-0.925
8	0.018	-0.068	-0.15	-0.051	-0.116
9	-0.69	-0.697	-0.698	-1.069	-1.193
10	0.097	-0.316	-0.383	-0.178	-0.227
11	-0.146	-0.316	-0.343	-0.126	-0.123
12	-0.869	-0.762	-0.753	-1.123	-1.175
13	-0.863	-0.975	-0.984	-0.838	-0.836
14	-0.881	-0.607	-0.483	-0.462	-0.259
15	-0.322	-0.225	-0.194	-0.147	-0.091
16	1.046	0.211	0.101	0.803	0.658
17	1	0.919	0.867	1.071	1.12
18	1.046	1.062	1.067	0.965	0.94
19	1.301	0.75	0.622	0.657	0.501
20	0.523	0.767	0.832	0.582	0.6
21	-0.934	-1.107	-1.127	-1.091	-1.109
22	-0.556	0.066	0.134	0.053	0.12
23	-1.037	-0.29	-0.081	-0.291	-0.082
24	-1.617	-1.089	-1.002	-0.8	-0.59
25	-0.041	0.1	0.134	0.167	0.218
26	0.398	-0.328	-0.447	-0.14	-0.248
27	0.699	0.126	-0.008	0.299	0.192
28	1.000	*	*	*	*
29	-1.312	-1.238	-1.227	-1.554	-1.616
30	-0.477	-0.039	0.014	-0.354	-0.327
31	-0.919	-0.444	-0.369	-0.752	-0.708
32	-0.653	0.211	0.326	0.016	0.129
33	-1.389	*	*	*	*

^a All data represent mean values for at least two separate experiments. * - Outliers, Obs. Act. - Observed activity, Cal. Act. - Calculated activity, Pred. Act. - Predicted activity by leave one out cross validation.

Consequently equation-4 can be considered as most suitable model with both high statistical significant and excellent predictive ability. The predictive ability of model-4 was also confirmed by external r^2CV_{ext} . The r^2CV_{ext} values are shown in Table 3. The r^2CV_{ext} value of the selected model is greater than the prescribed value ($r^2CV_{ext} > 0.5$). The robustness of the selected model was checked by Y – randomization test. The low r^2 and q^2 values indicate (data not shown) that the good results in our original model are not due to a chance correlation or structural dependency of the training set. The predictive ability of this model was also confirmed by leave 20 % out cross validation (equation 5).

$$pEC_{50} = 13.99 (\pm 9.130) + 0.302 (\pm 0.139) HP + 4.34 (\pm 1.152) HD + 0.051 (\pm 0.016) MR - 2.39 (\pm 1.103) IP + 0.898 (\pm 0.249) IH \quad (5)$$

Number of cycle 5 [compounds were deleted in 5 cycles in the following manner: (1, 6, 11....), (2, 7, 12.....), (3, 8, 13....), (4, 9, 14.....) and (5, 10, 15...)], $q^2 = 0.608$, Average of absolute values of predicted residual = 0.157.

Table 3. Prediction of anti-HIV activity of test set compounds in 5 cross validation cycles (leave 20 %-out) based on the descriptor set of equation 4

Cycle	Test set compounds	Training set compounds	Regression coefficients						
			Intercept	HP	HD	MR	IP	IH	r^2CV_{ext}
1	1, 6, 11, 16, 21, 26	Rest of the compounds (n = 24)	14.595	0.258	4.234	0.046	-2.373	0.980	0.607
2	2, 7, 12, 17, 22, 27	Rest of the compounds (n = 24)	11.404	0.284	3.897	0.046	-2.045	1.023	0.691
3	3, 8, 13, 18, 23, 28	Rest of the compounds (n = 24)	13.446	0.245	4.935	0.0628	-2.413	0.875	0.552
4	4, 9, 14, 19, 24, 29	Rest of the compounds (n = 24)	22.517	0.421	4.475	0.047	-3.359	0.868	0.618
5	5, 10, 15, 20, 25, 30	Rest of the compounds (n = 24)	8.015	0.301	4.135	0.0536	-1.782	0.742	0.573

$$r^2CV_{ext} = 1 - \frac{\sum_{i=1}^{test} (y_{exp} - y_{pred})^2}{\sum_{i=1}^{test} (y_{exp} - \bar{y}_{tr})^2}$$

Where \bar{y}_{tr} is the averaged value for the dependent variable for the training set.

The variables used in the selected model have no mutual correlation. This model showed good correlation coefficient (r) of 0.864 between descriptors [electronic (ionization potential), hydrogen bond donor, Hansen-polarity, indicator variable (IH) and thermodynamic (molar refractivity)] and anti-HIV activity. Squared correlation coefficient (r^2) of 0.747 explains 74.7% variance in biological activity. The positive contribution of hydrogen bond donor and Hansen-polarity on the biological activity showed that the increase in the values of these parameters lead to better anti-HIV TTDs compounds. The positive coefficient of MR showed that the volume is responsible for the activity. The negative coefficient of IP indicated that the increase of IP is detrimental to biological activity and the positive coefficient of indicator variable IH (having value 1 if *meta* halogenated benzyl ring is present at R₁ position) showed that the *meta* substituted benzyl group at R₁ position of TTDs is conducive to activity. Based on the developed QSAR model, new anti-HIV TTDs derivatives can be designed with caution.

Experimental Section

General Procedures. Win CAChe 6.1 (molecular modeling software, a product of Fujitsu private limited, Japan), Molecular modeling pro 6.1.0 (trial version, Cambridge software Corp.), STATISTICA version 6 (StatSoft, Inc., Tulsa, USA).

Optimization of molecules structure

All 33 TTDs structures were built on workspace of Win CAChe 6.1 (molecular modeling software, a product of Fujitsu private limited, Japan) and energy minimization of the molecules was done using Allinger's MM2 force field followed by semi empirical PM3 method available in MOPAC module until the root mean square gradient value becomes smaller than 0.001 kcal/mol Å. Most stable structure for each compound was generated and used for calculating various physico-chemical descriptors like thermodynamic, steric and electronic values of descriptors.

Descriptors calculation, QSAR models development and validation

All the calculated descriptors (50 descriptors calculated by Win CAChe 6.1 and Molecular modeling pro 6.1.0, and 3 indicator variables, the complete descriptors data set of all compounds will be provided on request) were considered as independent variable and biological activity as dependent variable. STATISTICA version 6 (StatSoft, Inc., Tulsa, USA) software was used to generate QSAR models by stepwise multiple linear regression analysis. Statistical measures used were n-number of compounds in regression, r-correlation coefficient, r^2 -squared correlation coefficient, F- test (Fischer's value) for statistical significance, SEE- standard error of estimation, q^2 - cross validated correlation coefficient and correlation matrix to show correlation among the parameters.

The squared correlation coefficient (or coefficient of multiple determination) r^2 is a relative measure of fit by the regression equation. Correspondingly, it represents the part of the variation

in the observed data that is explained by the regression. The correlation coefficient values closer to 1.0 represent the better fit of the regression. The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F-test indicate that the model is statistically significant. Standard deviation is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus standard deviation is an absolute measure of quality of fit and should have a low value for the regression to be significant.

The predictive ability of the generated correlations was evaluated by cross validation method employing a 'leave-one-out' scheme. Validation parameters considered were cross validated r^2 or q^2 , standard deviation based on predicted residual sum of squares (S_{PRESS}) and standard error of prediction (SDEP). The predictive ability of the selected model was also confirmed by external $r^2\text{CV}_{\text{ext}}$.

$$r^2\text{CV}_{\text{ext}} = 1 - \frac{\sum_{i=1}^{\text{test}} (y_{\text{exp}} - y_{\text{pred}})^2}{\sum_{i=1}^{\text{test}} (y_{\text{exp}} - \bar{y}_{\text{tr}})^2}$$

Where \bar{y}_{tr} is the averaged value for the dependent variable for the training set.

The robustness of a QSAR model was checked by Y – randomization test. In this technique, new QSAR models were developed by shuffling the dependent variable vector randomly and keeping the original independent variable as such. The new QSAR models are expected to have low r^2 and q^2 values. If the opposite happens then an acceptable QSAR model can not be obtained for the specific modeling method and data.

Acknowledgments

One of the authors V. Ravichandran is thankful to AICTE, New Delhi for providing (QIP) Senior Research Fellowship.

References

1. Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B. *Science* **1984**, *224*, 500.
2. Barre-Sinoussi, F.; Chermann, J. C.; Rey, R.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, w.; Montagnier, L. *Science* **1983**, *220*, 868.
3. De Clercq, E. *J. Med. Chem.* **1995**, *38*, 2491.
4. Milton, J.; Slater, M. J.; Bird, A. J.; Spinks, D.; Scott, G.; Price, C. E.; Downing, S.; Green, D. V. S.; Madar, S.; Bethell, R.; Stammers, D. K. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2623.

5. Young, S. D.; Britcher, S. F.; Tran, L. O.; Payne, L. S.; Lumma, W. C.; Lyle, T. A.; Huff, J. R.; Anderson, P. S.; Olsen, D. B.; Carroll, S. S. *Antimicrob. Agents Chemother.* **1995**, *39*, 2602.
6. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
7. Bohm, M.; Sturzebecher, J.; Klebe, G.; *J. Med. Chem.* **1999**, *42*, 458.
8. Pungpo, P.; Hannongbua, S. *J. Mol. Graphics & Modell.* **2000**, *18*, 581.
9. Ravichandran, V.; Agrawal, R. K. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2197.
10. Barreca, M. L.; Carotti, A.; Chimirri, A.; Monforte, A. M. *Bioorg. Med. Chem.* **1999**, *7*, 2283.
11. Nair, A. C.; Jayatilleke, P.; Wang, X.; Miertus, S.; Welsh, W. J. *J. Med. Chem.* **2002**, *45*, 973.
12. Jayatilleke, P. R. N.; Nair, A. C.; Zauhar, R.; Welsh, W. J. *J. Med. Chem.* **2000**, *43*, 4446.
13. Hilgeroth, A.; Fleischer, R.; Wiese, M.; Heinemann, F. W. *J. Computer-Aided Mol. Design* **1999**, *13*, 233.
14. A. K. Debnath, *J. Med. Chem.* **1999**, *42*, 249.
15. Oprea, T. I.; Waller, C. L.; Marshall, G. R. *Drug Design Discovery* **1994**, *12*, 29.
16. Buolamwini, J. K.; Assefa, H. *J. Med. Chem.* **2002**, *45*, 841.
17. Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W. *J. Med. Chem.* **1995**, *38*, 890.
18. Sahu, K. K.; Ravichandran, V.; Mourya, V. K.; Agrawal, R. K. *Med. Chem. Res.* **2007** (In press).
19. Debnath, A. K.; Jiang, S.; Strick, N.; Lin, K.; Haberfield, P. *J. Med. Chem.* **1994**, *37*, 1099.
20. Arranz, M. E.; Diaz, J. A.; Ingate, S. T.; Witvrouw, M.; Pannecouque, C.; Balzarini, J.; Clercq, E. D.; Vega, S. *Bioorg. Med. Chem.* **1999**, *7*, 2811.