Journal of Computer-Aided Molecular Design, **13:** 563–577, 1999.
KLUWER/ESCOM

Quantitative structure-activity relationships and comparative molecular field analysis of TIBO derivatised HIV-1 reverse transcriptase inhibitors

Supa Hannongbua[a,*], Pornpan Pungpo[a], Jumras Limtrakul[a] & Peter Wolschann[b]
[a]*Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand;* [b]*Institut für Theoretische Chemie und Strahlenchemie der Universität Wien, Währinger Strasse 17, A-1090 Wien, Austria*

Received 11 September 1998; Accepted 8 March 1999

Key words: HIV-1, molecular modeling, NNRTIs, QSAR, quantum chemical calculations, TIBO derivatives, 3D-QSAR

Summary

Quantitative structure-activity relationships (QSAR) and Comparative Molecular Field Analysis (CoMFA) have been applied in order to explain the structural requirements of HIV-1 reverse transcriptase (HIV-1 RT) inhibitory activity of TIBO derivatives on the MT-4 cells. The best QSAR model is satisfactory in both statistical significance and predictive ability. The derived structural descriptors indicate the importance of electronic contributions toward the HIV-1 RT inhibition of this class of compounds. However, it could not reveal any hydrophobic influence because of high collinearity between C2 and log *P* variables. In order to cope with steric interaction in the correlation, 3D-QSAR was performed using CoMFA. The obtained CoMFA model shows high predictive ability, $r^2_{cv} = 0.771$, and clearly demonstrates its potential in the steric feature of the molecules through contour maps, explaining a majority (81.8%) of the variance in the data. Consequently, these results can be useful in identifying the structural requirements of TIBO derivatives and helpful for better understanding the HIV-1 RT inhibition. Eventually, they provide a beneficial basis to design new and more potent inhibitors of HIV-1 RT.

Introduction

TIBO or tetrahydroimidazo-[4, 5, 1-jk][1, 4]-benzodiazepinone (Figure 1), one of the most specific and potent nonnucleoside reverse transcriptase inhibitors (NNRTI) of human immunodeficiency virus type 1 (HIV-1) replication, and its derivatives were developed by Pauwels et al. [1]. The available kinetic [2–4] and structural [5–7] studies make it clear that NNRTIs inhibit the enzyme noncompetitively and they all bind at a common allosteric site in HIV-1 RT. However, the effectivity is reduced by the very rapid development of drug-resistance mutations [8]. This causes a limitation to the potential utilization of this inhibitor. Numerous experimental investigations on complex structures of HIV-1 RT with TIBO and other NNRTIs [9, 10] have been performed, however the mechanism of action of this inhibitor as well as of drug resistance has still not been clarified. During complexation of HIV-1 RT with NNRTIs, the side chain and backbone of residues surrounding the pocket adjust to each bound drug in a common fashion. These results reveal that this protein is able to accommodate inhibitors of different chemical structures. The flexibility of the HIV-1 RT binding pocket implies that it may be not feasible to generate a generic NNRT binding site that could be used to model RT in the search for more potent drugs.

Our previous studies on HIV-1 RT inhibitors indicated that electronic and molecular properties of the inhibitors obtained from quantum calculations can be used as the structural descriptors [11, 12]. By considering the common features observed with these chemically divergent NNRTI, we have attempted to investigate the functional roles of particular groups. In order to get a closer insight into the structural requirements for a powerful inhibitor, we have extended our studies to include a series of compounds that show

*To whom correspondence should be addressed. E-mail: fscisph@ku.ac.th; Tel. 066-2-9428034; Fax. 066-2-5793955.

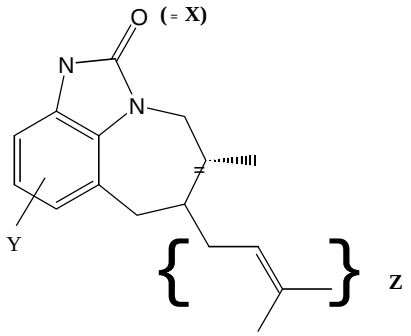

Figure 1. General structure of tetrahydroimidazo [4,5,1-jk][1, 4] benzodiazepinone.

widely different potencies. The compounds studied belong to the TIBO chemical series (Table 1).

In an attempt to cope with such problems, it is required to have a precise and detailed understanding of the important structure-activity relationships. Therefore, a quantitative structure-activity relationship (QSAR) study [13] was performed. In this procedure, atomic net charges obtained from molecular orbital calculations were used as the electronic descriptor. In addition, lipophilicity (log *P*), molar refractivity (MR) and molar polarizability (POL), were used as molecular properties. In order to use a more general approach to deal with steric interaction, Comparative Molecular Field Analysis (CoMFA), developed by Cramer et al. [14], was employed. This analysis aims to establish a relationship between HIV-1 RT inhibition of TIBO analogues and steric and electrostatic fields around them.

In the present work, QSAR and CoMFA have been performed for the first line with aims of (a) determining quantitative structure-activity relationship and structural requirements of HIV-1 NNRTI in the class of TIBO derivatives, (b) obtaining information about the structural characteristics underlying the inhibition of this class of compounds.

Methods of calculation

Biological data

The chemical structures of 46 TIBO derivatives are illustrated in Table 1, together with their biological activities, expressed as log(1/C), where C is the effective concentration of a compound required to achieve 50% (IC_{50}) protection of MT-4 cell against the cytopathic effect of HIV-1. The log(1/C) value was used as the dependent variable in SAR analysis. In CoMFA, these 46 TIBO derivatives are used as training set. Addition of 24 compounds (compounds T1-T24) was done in order to test the validation of the model (test set, see Table 8). The activity data of these inhibitors was taken from References 15, 28 and 30.

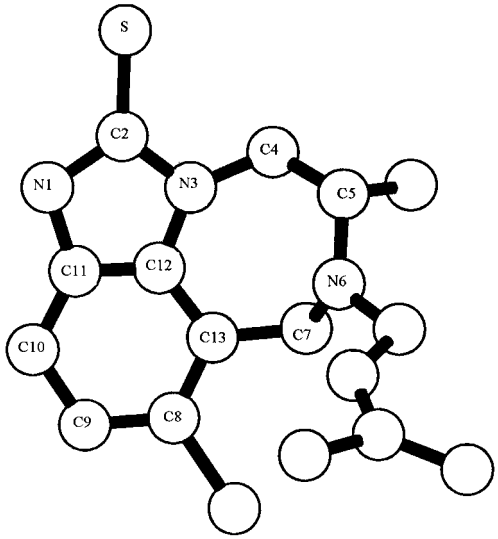

Figure 2. Structure of TIBO, obtained from complexed structure between TIBO and HIV-1 reverse transcriptase [9]; the atomic numbering as used in this study is indicated.

Calculations of structural properties for QSAR analysis

The starting geometry of TIBO (Figure 2) is obtained from a crystallographic structure of the enzyme-inhibitor complex [9]. Modification of substituents of all TIBO derivatives was done by ALCHEMY III [16]. Full geometry optimization of all structures was carried out with the AM1 semiempirical molecular orbital method [17], implemented in the GAUSSIAN 94 program [18] on a DEC AlphaStation model 250/4 266. Consequently, partial atomic charges were derived. Based on optimized geometry, molecular properties were calculated by the Chemplus 1.0 program [19]. Molecular polarizability (POL) can be expressed as the sum of atomic polarizabilities, plus corrections depending on the types of bonds present. These have been presented as approximate sums of bond polarizabilities [20]. The other investigated molecular descriptors, MR and log *P*, were determined by the same procedure as described in the previous study [11]. The multiple linear regression (MLR) in SPSS for Windows Release 6.0 [21] was used to develop a

Table 1. Structure of TIBO derivatives[a], experimental and calculated log (1/*C*) HIV-1 RT inhibitory affinities

Compd. No	*R*	*X*	*Y*	*Z*	log(1/*C*)		Residual
					Experimental	Calculated[b]	
1	H	S	8-Cl	DMA[c]	7.340	8.006	−0.666
2	H	S	9-Cl	DMA	6.790	7.167	−0.377
3	5-CH2CH3	O	–	2-MA[d]	4.300	4.958	−0.658
4	5-$CH(CH3)_2$	O	–	2-MA	5.000	5.051	−0.051
5	5-$CH(CH3)_2$	O	–	DMA	5.000	5.044	−0.044
6	5,5-di-CH3	O	–	2-MA	4.640	4.994	−0.354
7	4-CH3	O	–	2-MA	4.490	4.521	−0.031
8	4-CH3(S)	S	9-Cl	2-MA	6.170	6.578	−0.408
9	4-CH3	S	9-Cl	$CH2CH(CH2)_2$	5.660	6.416	−0.756
10	4-$CH(CH3)_2$	O	–	CH2CH2CH3	4.130	4.433	−0.303
11	4-$CH(CH3)_2$	O	–	2-MA	4.900	4.419	0.481
12	4-CH2CH2CH3	O	–	CH2CH2CH3	3.740	3.598	0.142
13	4-CH2CH2CH3	O	–	2-MA	4.320	4.469	−0.149
14	7-CH3	O	–	CH2CH2CH3	4.080	5.021	−0.941
15	7-CH3	O	–	DMA	4.920	5.016	−0.096
16	7-CH3	O	8-Cl	DMA	6.840	6.671	0.169
17	7-CH3	O	9-Cl	DMA	6.790	5.787	1.003
18	7-CH3	S	–	CH2CH2CH3	5.610	6.260	−0.650
19	7-CH3	S	–	DMA	7.110	6.266	0.845
20	7-CH3	S	8-Cl	DMA	7.920	7.932	−0.012
21	7-CH3	S	9-Cl	DMA	7.640	7.115	0.525
22	4,5-di-CH3(cis)	O	–	DMA	4.250	4.346	−0.096
23	4,5-di-CH3(cis)	S	–	DMA	5.650	5.727	−0.077
24	4,5-di-CH3(trans)	S	–	$CH2CH(CH2)_2$	4.870	4.725	0.145
25	4,5-di-CH3(trans)	S	–	DMA	4.840	5.753	−0.913
26	4-keto-5-CH3	S	9-Cl	CH2CH2CH3	4.300	3.959	0.341
27	4,5-benzo	S	–	$CH2CH(CH2)_2$	5.000	4.944	0.056
28	5,7-di-CH3(trans)	S	–	DMA	7.380	6.193	1.187
29	5,7-di-CH3(cis)	S	–	DMA	5.940	6.265	−0.325
30	5,7-di-CH3(R,R; trans)	O	9-Cl	DMA	6.640	5.771	0.869
31	5,7-di-CH3 (R,R; trans)	S	9-Cl	DMA	6.320	7.012	−0.692
32	5,7-di-CH3 (S,S; trans)	O	9-Cl	DMA	5.300	5.127	0.173
33	4,7-di-CH3(trans)	S	–	DMA	4.590	4.423	0.167
34	5,6-CH2C(=CHCH3)CH2(S)	S	9-Cl	–	5.420	6.126	−0.706
35	6,7-(CH2)4	S	9-Cl	–	5.700	6.934	−1.234
36	5-CH3(S)	S	8-Cl	DMA	8.300	7.880	0.420
37	5-CH3(S)	O	9-Cl	DMA	6.740	5.840	0.900
38	5-CH3(S)	S	9-Cl	DMA	7.370	7.066	0.304
39	5-CH3(S)	S	9-Cl	$CH2CH(CH2)_2$	7.470	7.079	0.391
40	5-CH3(S)	S	–	$CH2CH(CH2)_2$	7.220	6.261	0.959
41	5-CH3	O	–	CH2CH2CH3	4.220	5.127	−0.907
42	5-CH3	S	–	CH2CH2CH3	5.780	6.285	−0.505
43	5-CH3	O	–	2-MA	4.460	5.071	−0.611
44	5-CH3	S	–	DMA	7.010	6.281	0.729
45	5-CH3(S)	O	–	DMA	5.480	5.054	0.426
46	5-CH3(S)	S	–	2-MA	7.580	6.237	1.344

[a] See Figure 1.
[b] Calculated by Equation 5.
[c] DMA = 3,3-dimethylallyl.
[d] 2-MA = 2-methylallyl.

QSAR model. The validity of the model was proven by multiple correlation coefficients (r), standard deviation (s) and F-test value. The reliability of the model was indicated in terms of predictive $r^2(Q^2)$.

Alignment rule and CoMFA analysis

The alignment rule, i.e., the positioning of a molecular model within a fixed lattice, is the most important input variable in CoMFA. In this study, all TIBO structures were fully optimized by ab initio calculations with the HF/3-21G basis set. The availability of crystallographic data for the HIV-1 RT/TIBO complex structure offered a possibility for alignment rule, i.e., minimization within the active site. To reduce the computational complexity, a substructure sphere of 12 Å radius centered on the ligand of the inhibitor/enzyme complex was defined as active site. The rigid fragments common to investigated compounds were superimposed to the equivalent atoms in a TIBO template molecule. Backbone atoms of the selected active site were kept rigid during minimization. Side chain atoms and ligand atoms were allowed to relax. After minimization, these ligands were extracted from the active site. For the active site, partial atomic charges were loaded from the SYBYL Biopolymer dictionary (Kollman all atom method). Partial atomic charges required for calculations of electrostatic interaction were computed by the MOPAC program [22] in SYBYL 6.4 [23], using the semiempirical method AM1.

A CoMFA cubic lattice, with 2 Å grid spacing, was generated around these molecules based on the molecular volume of the structures. These dimensions ensured that the grid extended beyond the molecular dimensions by 4.0 Å in all directions. In this investigation, three different atoms, sp^3 carbon atom with +1 charge (default probe atom in SYBYL), sp^3 oxygen atom with -1 charge and H atom with $+1$ charge, served as probe atoms. The probe atom was placed at each lattice point and the interactions of the steric and electrostatic fields with each atom in the molecule were all calculated with CoMFA standard scaling and then put in a CoMFA QSAR table. In order to speed up the analysis and reduce the amount of noise, the minimum sigma value was set to 2.0 kcal/mol and energy cutoff values of 30 kcal/mol were selected for both electrostatic and steric fields.

For the CoMFA model, partial least-squares (PLS) methodology was employed to perform the correlation between the steric and electrostatic properties and inhibitory activity. The orthogonal latent variables were extracted by the NIPALS algorithm [24] and subjected to full cross-validation (leave-one-out method). The analyses were carried out with a maximum of ten components, and subsequently, using the number of component (noc) at which the difference in the r^2_{cv} value to the next one was less than 0.02 [25]. Following the cross-validated analysis, a non cross-validated analysis was performed using the optimal number of components previously identified and was then employed to analyze the CoMFA results.

Predictive ability

The overall predictive ability of the analysis was expressed in terms of Q^2 or r^2_{cv}, in QSAR and CoMFA, respectively, which is defined as

$$Q^2 \text{ or } r^2_{cv} = (\text{SSY} - \text{PRESS})/\text{SSY},$$

where SSY represents the variance of the biological activities of molecules around the mean value and PRESS is the prediction error sum of squares derived from the leave-one-out method. The uncertainty of the prediction is defined as

$$S_{\text{PRESS}} = [\text{PRESS}/(n - k - 1)]^{1/2},$$

where k is the number of variables in the model and n is the number of compounds used in the study.

Results

QSAR analysis

The following atomic net charges of various atoms in tricyclic rings, N1, C2, N3, C4, C5, N6, C7, C8, C9, C10, C11, C12 and C13, illustrated in Figure 1, were considered as electronic variables. In order to account for molecular properties of molecules, partition coefficients (log P), molecular refractivity (MR) and molecular polarizability (POL) were employed. All investigated variables are presented in Table 2.

All possible combinations of parameters were considered to develop the QSAR model. A large number of satisfactory statistical models was obtained. The quality of models, judged by r, s, F, Q^2 and S_{PRESS}, was statistically tested. Consequently, Equations 1–9 were consecutively evaluated. Based on statistics for the coefficients of variables of Equations 7–9 (Table 3) and possible chance correlations [26] (Table 4), the use of more than six independent variables was not

Table 2. Atomic net charge and molecular properties for compounds in Table 1, used in the QSAR study

No.	log *P*	MR	POL	N1	C2	N3	C4	C5	N6	C7	C8	C9	C10	C11	C12	C13
1	3.630	56.453	34.436	−0.4068	0.0720	−0.2911	−0.1451	−0.1446	−0.3371	−0.0977	−0.0777	−0.1650	−0.1795	0.0234	0.0272	−0.0790
2	3.630	56.453	34.436	−0.4074	0.0745	−0.2939	−0.1421	−0.1445	−0.3383	−0.0911	−0.1722	−0.0701	−0.1745	0.0318	0.0225	−0.0917
3	4.854	48.804	31.183	−0.4628	0.4604	−0.3211	−0.1358	−0.0440	−0.3452	−0.0881	−0.1871	−0.1775	−0.1932	0.0374	0.0107	−0.0965
4	5.257	53.276	33.018	−0.4390	0.4508	−0.3274	−0.1318	−0.0362	−0.3470	−0.0854	−0.1824	−0.1812	−0.1862	0.0203	0.0162	−0.0991
5	5.571	57.884	34.853	−0.4324	0.4514	−0.3373	−0.1226	−0.0257	−0.3429	−0.0858	−0.1812	−0.1837	−0.1841	0.0137	0.0302	−0.1032
6	4.464	48.917	31.183	−0.4342	0.4546	−0.3406	−0.1216	0.0488	−0.3349	−0.0892	−0.1826	−0.1830	−0.1848	0.0159	0.0289	−0.1020
7	4.386	44.280	29.348	−0.4457	0.4457	−0.3010	−0.0599	−0.1125	−0.3684	−0.0781	−0.1867	−0.1785	−0.1893	0.0253	0.0086	−0.0971
8	3.730	56.263	34.436	−0.4074	0.0661	−0.2728	−0.0605	−0.1119	−0.3731	−0.0771	−0.1740	−0.0690	−0.1765	0.0316	0.0127	−0.0866
9	3.211	61.899	33.854	−0.4073	0.0749	−0.2876	−0.0437	−0.1292	−0.3487	−0.0850	−0.1779	−0.0675	−0.1777	0.0318	0.0246	−0.0892
10	4.840	56.242	31.375	−0.4388	0.4470	−0.3056	−0.0430	−0.0924	−0.3924	−0.0487	−0.1864	−0.1797	−0.1869	0.0174	0.0126	−0.0994
11	5.257	53.276	33.018	−0.4358	0.4495	−0.3140	−0.0397	−0.1055	−0.3823	−0.0706	−0.1851	−0.1810	−0.1859	0.0149	0.0176	−0.0997
12	4.834	56.372	31.375	−0.4428	0.4443	−0.3058	−0.0500	−0.1019	−0.3928	−0.0610	−0.1888	−0.3058	−0.1895	0.0223	0.0136	−0.0968
13	5.251	53.405	33.018	−0.4413	0.4454	−0.3001	−0.0556	−0.1138	−0.3677	−0.0777	−0.1857	−0.1787	−0.1885	0.0218	0.0041	−0.0939
14	3.969	47.246	27.705	−0.4446	0.4485	−0.3154	−0.1444	−0.1422	−0.3385	0.0051	−0.1865	−0.1789	−0.1883	0.0245	0.0149	−0.0938
15	4.700	48.888	31.183	−0.4409	0.4475	−0.3175	−0.1444	−0.1378	−0.3421	0.0033	−0.1847	−0.1797	−0.1878	0.0217	0.0135	−0.0910
16	5.328	53.382	33.111	−0.4372	0.4465	−0.3081	−0.1509	−0.1372	−0.3430	−0.0017	−0.0778	−0.1664	−0.1829	0.0236	0.0171	−0.0694
17	5.328	53.382	33.111	−0.4517	0.4541	−0.3153	−0.1456	−0.1380	−0.3429	0.0038	−0.1735	−0.0698	−0.1804	0.0395	0.0128	−0.0818
18	2.685	54.735	30.865	−0.4077	0.0743	−0.2916	−0.1424	−0.1427	−0.3408	0.0048	−0.1857	−0.1779	−0.1857	0.0213	0.0226	−0.0930
19	3.419	56.377	34.343	−0.4080	0.0731	−0.2891	−0.1435	−0.1424	−0.3354	0.0031	−0.1859	−0.1778	−0.1859	0.0215	0.0208	−0.0927
20	4.044	60.871	36.271	−0.4079	0.0671	−0.2827	−0.1489	−0.1430	−0.3340	−0.0029	−0.0790	−0.1645	−0.1809	0.0238	0.0240	−0.0712
21	4.044	60.871	36.271	−0.4075	0.0616	−0.2761	−0.1557	−0.1318	−0.3513	0.0084	−0.1706	−0.0707	−0.1756	0.0314	0.0027	−0.0751
22	5.113	53.306	33.018	−0.4483	0.4499	−0.3043	−0.0399	−0.0424	−0.3357	−0.0908	−0.1890	−0.1769	−0.1913	0.0271	0.0102	−0.0968
23	3.829	60.795	36.178	−0.4077	0.0667	−0.2728	−0.0536	−0.0164	−0.3790	−0.0694	−0.1860	−0.1770	−0.1871	0.0211	0.0095	−0.0944
24	2.996	61.823	33.761	−0.4097	0.0766	−0.2798	−0.0028	−0.0449	−0.3356	−0.0874	−0.1907	−0.1771	−0.1885	0.0206	0.0271	−0.0490
25	3.829	60.795	36.178	−0.4074	0.0638	−0.2652	−0.0553	−0.0159	−0.3704	−0.0719	−0.1861	−0.1772	−0.1868	0.0205	0.0104	−0.0950
26	5.062	56.463	32.879	−0.4185	0.0679	−0.3371	0.3413	−0.0431	−0.3492	−0.0883	−0.1786	−0.0615	−0.1833	0.0525	0.0219	−0.0891
27	3.844	47.138	36.081	−0.4068	0.0871	−0.2571	0.0668	0.0508	−0.3603	−0.0615	−0.1845	−0.1758	−0.1850	0.0235	0.0247	−0.0974
28	3.829	60.795	36.178	−0.4069	0.0792	−0.3000	−0.1342	−0.0478	−0.3418	0.0048	−0.1842	−0.1788	−0.1858	0.0198	0.0257	−0.0916
29	3.829	60.795	36.178	−0.4072	0.0659	−0.2806	−0.1482	−0.0387	−0.3498	0.0112	−0.1824	−0.1793	−0.1856	0.0204	0.0050	−0.0843
30	5.741	57.801	34.946	−0.4476	0.4548	−0.3222	−0.1375	−0.0474	−0.3407	0.0066	−0.1724	−0.0712	−0.1789	0.0359	0.0179	−0.0841
31	4.457	65.298	38.106	−0.4067	0.0774	−0.2996	−0.1352	−0.0476	−0.3425	0.0054	−0.1718	−0.0709	−0.1754	0.0297	0.0270	−0.0835
32	5.741	57.801	34.946	−0.4369	0.4423	−0.2954	−0.1536	−0.0284	−0.3546	0.0106	−0.1722	−0.0776	0.0315	−0.0033	−0.0033	−0.0757
33	3.829	60.795	36.178	−0.4072	0.0640	−0.2679	−0.0624	−0.1092	−0.3718	0.0171	−0.1865	−0.1865	−0.1876	0.0201	0.0111	0.0111
34	3.216	54.391	33.662	−0.4071	0.0883	−0.3119	−0.0777	−0.0737	−0.3252	−0.0890	−0.1767	−0.0713	−0.1740	0.0271	0.0542	−0.0516
35	2.881	57.351	32.019	−0.4077	0.0682	−0.2844	−0.1523	−0.1251	−0.3626	0.0168	−0.1745	−0.0713	−0.1761	0.0306	0.0116	−0.0694
36	4.044	60.871	36.271	−0.4055	0.0790	−0.3014	−0.1348	−0.0438	−0.3433	−0.0954	−0.0774	−0.1661	−0.1799	0.0209	0.0344	−0.0763
37	5.328	53.382	33.111	−0.4454	0.4553	−0.3423	−0.1360	−0.0420	−0.3423	−0.0883	−0.1714	−0.0714	−0.1779	0.0349	0.0174	−0.0895
38	4.044	60.871	36.271	−0.4064	0.0782	−0.2986	−0.1349	−0.0437	−0.3430	−0.0898	−0.1716	−0.0706	−0.1752	0.0301	0.0247	−0.0879
39	3.211	61.899	33.854	−0.4074	0.0753	−0.2893	−0.1414	−0.0290	−0.3686	−0.0726	−0.1754	−0.0702	−0.1759	0.0304	0.0222	−0.0897
40	2.583	57.405	31.926	−0.4075	0.0771	−0.2900	−0.1404	−0.0288	−0.3679	−0.0728	−0.1878	−0.1782	−0.1863	0.0204	0.0209	−0.0978
41	3.969	47.246	27.705	−0.4263	0.4465	−0.3215	−0.1380	−0.0488	−0.3421	−0.0871	−0.1830	−0.1813	−0.1839	0.0149	0.0174	−0.1017
42	2.685	54.735	30.865	−0.4071	0.0809	−0.3022	−0.1333	−0.0496	−0.3434	−0.0875	−0.1843	−0.1784	−0.1848	0.0212	0.0252	−0.0999
43	4.386	44.280	29.348	−0.4390	0.4517	−0.3284	−0.1345	−0.0494	−0.3409	−0.0877	−0.1826	−0.1809	−0.1861	0.0208	0.0174	−0.0997
44	3.415	56.377	34.343	−0.4068	0.0816	−0.3036	−0.1328	−0.0488	−0.3434	−0.0878	−0.1830	−0.1787	−0.1850	0.0205	0.0247	−0.0984
45	4.700	48.888	31.183	−0.4375	0.4498	−0.3249	−0.1350	−0.0418	−0.3416	−0.0885	−0.1832	−0.1809	−0.1863	0.0193	0.0172	−0.0982
46	3.101	51.768	32.508	−0.4069	0.0806	−0.2997	−0.1336	−0.0444	−0.3410	−0.0914	−0.1844	−0.1785	−0.1857	0.0207	0.0237	−0.0959

Table 3. The 95% confidence, F and p statistics for the coefficients of variables in Equations 7–10

Equation	Variable	Coefficient	95% confidence	F	p	Significance
7	C8	13.009	±7.168	13.499	0.001	
	C2	−6.194	±2.424	26.759	0.000	
	C4	−8.104	±2.630	38.922	0.000	
	C9	4.302	±4.063	4.384	0.043	
	C13	−11.842	±10.958	4.786	0.035	
	log P	0.514	±0.413	6.376	0.016	
	N3	−8.550	±13.818	1.569	0.218	not significant
8	C8	13.828	±7.510	13.919	0.000	
	C2	−4.303	±5.427	2.581	0.117	not significant
	C4	−7.163	±3.580	16.438	0.000	
	C9	4.497	±4.156	4.806	0.034	
	C13	−12.375	±11.106	5.096	0.030	
	log P	0.161	±0.675	0.107	0.745	not significant
	N3	−10.742	±14.993	2.108	0.155	not significant
	POL	0.097	±0.055	0.624	0.435	
9	C8	14.277	±7.624	14.426	0.000	
	C2	−5.329	±5.986	3.259	0.079	not significant
	C4	−7.163	±3.598	16.309	0.000	
	C9	4.204	±4.172	4.052	0.052	
	C13	−12.792	±11.206	5.360	0.026	
	log P	0.056	±0.922	0.012	0.913	not significant
	N3	−11.811	±15.285	2.456	0.126	not significant
	POL	0.123	±0.012	0.933	0.340	
	N1	−16.501	±39.716	0.710	0.405	not significant
10	C8	20.296	±8.537	22.764	0.000	
	C9	9.036	±4.550	16.109	0.003	
	C4	−4.928	±3.044	10.705	0.002	
	log P	−0.417	±0.284	8.820	0.005	
	C13	−7.398	±13.414	1.242	0.271	not significant

acceptable. Accordingly, only the best of the one-term, two-term, three-term, four-term, five-term and six-term models were considered.

$$\log(1/C) = 22.423(\pm 10.774)\mathbf{C8} + 9.631(\pm 1.885),$$
$$(n = 46, r = 0.534, s = 1.077, F = 17.951,$$
$$Q^2 = 0.234, S_{\mathrm{PRESS}} = 1.115), \quad (1)$$

$$\log(1/C) = 19.698(\pm 9.288)\mathbf{C8} - 3.067(\pm 1.479)\mathbf{C2}$$
$$+ 9.890(\pm 1.1614),$$
$$(n = 46, r = 0.702, s = 0.918, F = 20.844,$$
$$Q^2 = 0.435, S_{\mathrm{PRESS}} = 0.958), \quad (2)$$

$$\log(1/C) = 15.677(\pm 7.904)\mathbf{C8} - 3.521(\pm 1.242)\mathbf{C2}$$
$$-6.422(\pm 2.851)\mathbf{C4} + 8.666(\pm 1.443),$$
$$(n = 46, r = 0.812, s = 0.760, F = 27.13,$$
$$Q^2 = 0.531, S_{\mathrm{PRESS}} = 0.872), \quad (3)$$

$$\log(1/C) = 15.281(\pm 7.210)\mathbf{C8} - 2.982(\pm 1.186)\mathbf{C2}$$
$$-6.587(\pm 2.602)\mathbf{C4} + 6.195(\pm 4.035)\mathbf{C9}$$
$$+9.382(\pm 1.396),$$
$$(n = 46, r = 0.851, s = 0.693, F = 26.926,$$
$$Q^2 = 0.638, S_{\mathrm{PRESS}} = 0.766), \quad (4)$$

Table 4. Squared correlation matrix for parameters used in QSAR study

	Log *P*	MR	POL	N1	C2	N3	C4	C5	N6	C7	C8	C9	C10	C11	C12	C13	log(1/*C*)
Log *P*	1.000	−0.225	−0.005	−0.779	0.776	−0.559	0.114	0.093	0.093	0.066	0.060	−0.063	0.238	0.008	−0.363	−0.152	−0.307
MR		1.000	0.775	0.600	−0.627	0.469	−0.005	−0.069	−0.025	0.236	0.201	0.308	0.133	0.077	0.099	0.382	0.474
POL			1.000	0.532	−0.560	0.505	0.034	0.109	−0.143	0.207	0.277	0.299	0.159	0.059	0.102	0.356	0.523
N1				1.000	−0.954	0.658	0.041	−0.004	−0.134	0.020	0.164	0.233	−0.053	−0.025	0.395	0.282	0.520
C2					1.000	−0.690	−0.128	0.050	0.163	−0.049	−0.141	−0.311	0.099	−0.189	−0.336	−0.320	−0.526
N3						1.000	0.015	−0.139	−0.167	0.261	0.070	0.004	0.060	−0.112	−0.185	0.366	0.247
C4							1.000	0.214	0.003	−0.283	−0.204	0.072	−0.124	0.371	0.045	0.030	−0.432
C5								1.000	−0.065	−0.446	−0.263	−0.111	0.109	−0.173	0.131	−0.238	−0.178
N6									1.000	0.017	0.026	0.004	−0.042	−0.079	0.086	−0.093	−0.091
C7										1.000	0.058	0.058	0.230	−0.056	−0.356	0.366	0.146
C8											1.000	0.070	0.052	0.024	0.236	0.223	0.534
C9												1.000	0.051	0.667	0.181	0.163	0.401
C10													1.000	−0.428	−0.284	0.115	0.036
C11														1.000	0.106	0.039	0.146
C12															1.000	0.066	0.281
C13																1.000	0.094
log(1/*C*)																	1.000

$$\log(1/C) = 16.743(\pm 7.022)\mathbf{C8} - 3.320(\pm 1.175)\mathbf{C2} - 6.506(\pm 2.489)\mathbf{C4} + 6.473(\pm 3.866)\mathbf{C9} - 12.312(\pm 11.227)\mathbf{C13} + 8.697(\pm 1.474),$$
$$(n = 46, r = 0.869, s = 0.662, F = 24.579, Q^2 = 0.677, S_{\mathrm{PRESS}} = 0.724), \quad (5)$$

$$\log(1/C) = 13.604(\pm 7.149)\mathbf{C8} - 5.415(\pm 2.085)\mathbf{C2} - 7.867(\pm 2.618)\mathbf{C4} + 5.054(\pm 3.848)\mathbf{C9} - 13.552(\pm 10.671)\mathbf{C13} + 0.491(\pm 0.413)\log P + 6.148(\pm 2.610),$$
$$(n = 46, r = 0.887, s = 0.626, F = 23.887, Q^2 = 0.712, S_{\mathrm{PRESS}} = 0.726). \quad (6)$$

In the above correlation, n is the number of compounds used to derive the model, F is overall F-statistics for the addition of each successive term, and values in parentheses are the 95% confidence limit of each coefficient. The statistics for the coefficients are summarized in Table 3. The correlation matrix for the variables is given in Table 4. In order to examine the predictive power of the model, a cross-validation test was performed on the data set.

CoMFA analysis

Initially, a critical step in the construction of the CoMFA model was the attainment of an alignment rule. All TIBO compounds were fully optimized using ab initio calculations with the HF/3-21G basis set and the corresponding minimum energy conformers were aligned as described before. The effect of the probe atom was investigated for the reason that CoMFA depends on the interaction energy between probe atoms and molecules. All obtained analyses included both field types, i.e. steric and electrostatic fields. In addition, separated analysis of only steric or electrostatic field types was also performed. CoMFA with a default setting probe (sp^3C) yielded a model with r^2_{cv} of 0.570 (model 1) as shown in Table 5. Regarding the other probe atom, sp^3O (−1), produced better r^2_{cv} (model 2, $r^2_{cv} = 0.594$) than those obtained from sp^3C and H (+1) (model 3, $r^2_{cv} = 0.544$). It is important to note that steric field type models indicate higher predictive ability than that obtained from electrostatic field type models as shown in models 1–3.

The atomic charge of the sp^3O probe atom was thus selected for the next investigation on the effect of lattice spacing. The results of grid-CoMFA tests are summarized in Table 6. The r^2_{cv} indicated that the grid spacing set to 2 Å was suitable. A decrease in grid spacing leads to increasing the number of interaction energy values; however, it also increases the noise in PLS analysis.

Discussion

QSAR model

According to the QSAR models listed above, it was found that all equations show statistical significance. The decrease of F-value on addition of more parameters into Equations 2 and 3, producing Equation 4, indicates that the suitable equation would be either Equation 2 or 3. However, the predictive ability (Q^2) was not acceptable as the Q^2 value should be greater than 0.6 [27]. Therefore, only three models, Equations 4, 5, and 6 were considered to be reasonable predictive models. A comparison of these three models shows that the predictive ability of Equation 6 (Q^2) is the highest (0.712). This equation contains C2 and log P variables which show high collinearity (0.776, see Table 4). In order to evaluate the reliability of both variables in the equation, C2 or log P were excluded from Equation 6. After taking out log P, the result is identical to Equation 5. By this procedure, omitting C2 produced Equation 10.

$$\log(1/C) = 20.296(\pm 8.597)\mathbf{C8} + 9.036(\pm 4.550)\mathbf{C9} - 4.928(\pm 3.044)\mathbf{C4} - 0.417(\pm 0.28)\log P - 7.398(\pm 13.414)\mathbf{C13} + 11.2315(\pm 2.126)$$
$$(n = 46, r = 0.797, s = 0.807, F = 13.903, Q^2 = 0.536, S_{\mathrm{PRESS}} = 0.910) \quad (10)$$

It is observed that the statistical criteria of Equation 10 were lower than that of Equation 5. In addition, such a model contained one insignificant coefficient of C13 as expressed by the confidence interval and F-statistics (Table 3).

In comparing the quality of models, Equation 5 and Equation 4, it is evident that Equation 5 shows higher predictive ability (0.677). Consequently, Equation 5 was considered to be the best QSAR model in the

Table 5. Summary of CoMFA models with 46 TIBO compounds at different probe atoms

Model	Probe atom	Field type	Noc	r^2_{cv}	s-press	r^{2a}	*s*	*F*	Outliers (residual)	Steric contb[b]
1	$sp^3C(+1)$	both	3	0.570	0.854	0.847	0.510	77.590	cpd.39(−1.197)	79.3
		st	3	0.541	0.883	0.822	0.549	64.863		
		el	2	0.369	1.023	0.731	0.668	58.468		
2	$sp^3O(-1)$	both	3	0.594	0.831	0.860	0.487	86.247	cpd.33(−1.005)	79.6
		st	3	0.579	0.846	0.832	0.535	69.179		
		el	2	0.369	1.023	0.731	0.668	58.468		
3	H(+1)	both	3	0.544	0.880	0.850	0.505	79.079	cpd.28(1.05)	74.7
		st	3	0.541	0.883	0.822	0.549	64.863		
		el	2	0.369	1.023	0.731	0.668	58.468		
4[c]	$sp^3O(-1.0)$	both	3	0.629	0.796	0.877	0.459	97.07	cpd.46(0.908)	79.5
		st	3	0.603	0.822	0.847	0.519	72.74		
		el	3	0.475	0.958	0.880	0.458	73.344		
5[d]	$sp^3O(-1.0)$	both	3	0.657	0.756	0.887	0.434	104.743	cpd.39(1.01)	79.7
		st	3	0.629	0.786	0.853	0.494	77.669		
		el	4	0.533	0.894	0.900	0.412	88.238		
6[e]	$sp^3O(-1.0)$	both	3	0.689	0.712	0.901	0.402	118.543	cpd.15(0.996)	80.2
		st	3	0.660	0.744	0.869	0.462	86.547		
		el	4	0.471	0.941	0.890	0.430	76.565		
7[f]	$sp^3O(-1.0)$	both	3	0.721	0.680	0.919	0.367	143.049	Cpd.40(0.901)	81.5
		st	3	0.700	0.705	0.897	0.413	110.53		
		el	4	0.469	0.951	0.899	0.414	82.425		
8[g]	$sp^3O(-1.0)$	both	3	0.771	0.612	0.941	0.312	195.028	Cpd.21(0.662)	81.8
		st	2	0.741	0.644	0.920	0.363	141.242		
		el	2	0.487	0.905	0.785	0.586	69.785		

[a]Conventional r^2.
[b]Steric contribution in %.
[c]Elimination of compd. N. 33 (remaining 45 compds in the training set).
[d]Elimination of compds N. 33 and 46 (remaining 44 compds in the training set).
[e]Elimination of compds N. 33, 46 and 39 (remaining 43 compds in the training set).
[f]Elimination of compds N. 33, 46, 39 and 15 (remaining 42 compds in the training set).
[g]Elimination of compds N. 33, 46, 39 15 and 40 (remaining 41 compds in the training set).

present study, both in statistical significance and predictive ability. The variables used in this equation have no mutual correlation as shown in Table 4. The multiple correlation coefficient of this equation is 0.869 and it accounts for 75.5% of the variation in the biological data. The experimental versus the calculated affinities obtained from Equation 5 are reported in Table 1 and plotted in Figure 3.

Inspection of the best QSAR model (Equation 5) apparently indicates the importance of electronic characteristic contribution to the HIV-1 RT inhibition of TIBO analogues. Among the atomic charge variables, five indicators, C8, C2, C4, C9 and C13 were found to be the main parameters of influence in the correlation. It can be seen from all QSAR equations (Equations 1–10) that atomic charge at the C8-position seems to be the major contributor to the affinity. The positive value of the coefficient for this term suggests that low electron density on the C8-position correlates with increased pharmacological activity. Therefore, the presence of a strong electron-withdrawing substituent at this position is required. This is also in agreement with the empirical observation that compounds substituted by 8-Cl and/or 9-Cl atoms in the benzene ring are more active than those without it [28]. Consequently, the influence of the C9 parameter on biological activity could be explained in the same manner. In contrast, the coefficient of C2 showing a negative sign suggests that the lower positive charge at such positions leads to a higher affinity. Substitution of a sulfur atom for an oxygen atom at the C2-position lowers the positive charge on the site. This was consistently found by

Table 6. The statistic results with lattice grid space by 1 Å and 2 Å

Probe atom	Grid (Å)	Noc	r^2_{cv}	s-press	$r^{2\,a}$	*s*	*F*	Steric contb[b]
$sp^3O(-1.0)$	1	5	0.585	0.868	0.958	0.273	182.805	77.6
	2	3	0.594	0.831	0.860	0.487	86.247	79.6

[a]Conventional r^2.
[b]Steric contribution in %.

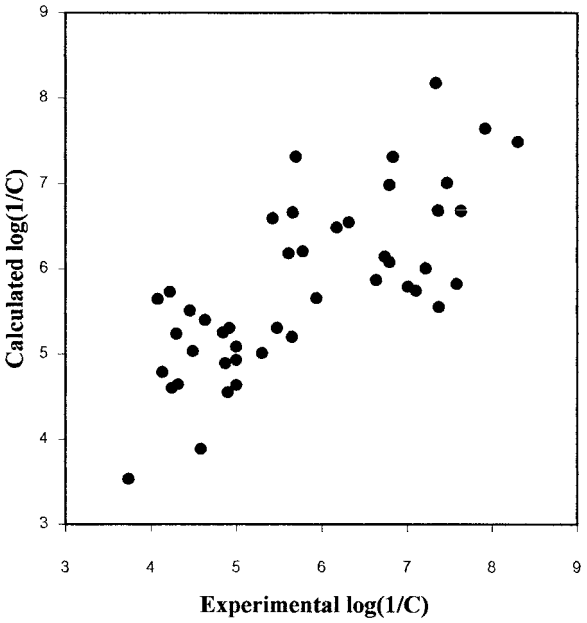

Figure 3. Comparison of experimental with calculated HIV-1 affinities (log(1/*C*)) obtained from conventional QSAR (Equation 5).

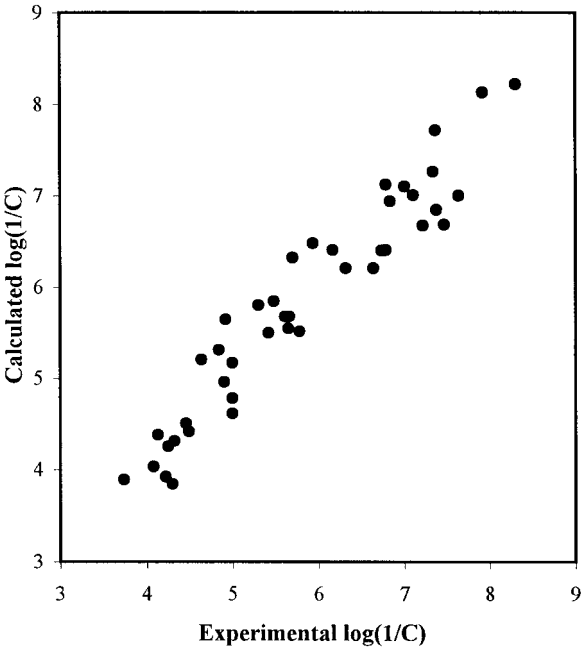

Figure 4. Plot of calculated versus experimental HIV-1 RT inhibitory affinities (log(1/*C*)) obtained from non-cross-validation of CoMFA model 8 for training set compounds.

Kukla et al. [29] who reported that the replacement of urea (one) with a thiourea (thione) at this position always yielded much more potent derivatives. It may be assumed that both sulfur and oxygen may be involved in some charge transfer interactions with the receptor. Considering the presence of the C4 atomic variable in the correlation, it indicates the importance of partial atomic charge on the carbon attached to the *R*-substituent. Its negative coefficients bring the desirable optimum electron density at this position. Due to there being a ring juncture between the benzene and diazepine rings of C13, the changes of substitution at several positions in cyclic rings also affect the variations of electron densities at each position, resulting in pharmacological activity.

Regarding the molecular parameters employed in the QSAR study, addition of a lipophilicity term, represented by log *P* to Equation 5, as presented in Equation 6, produced better statistical results. However, a high intercorrelation between log *P* and C2 (0.776, listed in Table 4) was encountered. Therefore, this model was found to be unreliable. It is also observed that other molecular properties, i.e. MR, a measure of substituent bulk, and POL were not accounted for in the obtained QSAR model. It could be explained by the high collinearity between both variables to each other and to C2.

It is interesting to note that atomic net charge of TIBO analogues, carried out by ab initio geometrical optimization with HF/3-21G basis sets, were also used as parameters in QSAR analysis. Nevertheless, we have consistently found that the quality of the QSAR models, obtained by ab initio (on the HF/3-21G level) charges, could not be improved, both in statistical

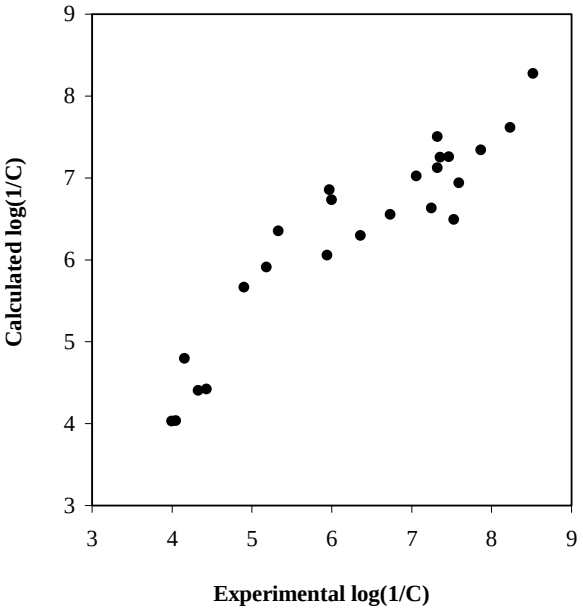

Figure 5. Plot of calculated versus experimental HIV-1 RT inhibitory affinities (log(1/*C*)) obtained from non-cross-validation of CoMFA model 8 for test set compounds.

significance and predictive ability, compared to those obtained from AM1 charges*.

CoMFA model

Model 2 expresses a higher level of internal consistency compared to the standard method (model 1), but one compound (compound 33) is significantly out of line as shown in Table 7. Inspection of outliers obtained from other CoMFA models was done and the maximum residual value is shown in Table 5. As compound 33 was found to be a maximum outlier, therefore, elimination of this compound from the CoMFA analysis produced model 4. However, a residual is still in the same range as obtained by compound 33. The consecutive outliers, compounds 46, 39, 15 and 40, were then removed, resulting in models 5, 6, 7 and 8, respectively. The final model, model 8, satisfied both predictive ability ($r^2_{cv} = 0.771$) and maximum outlier. The differences seen in connection with these outliers cannot be explained by the model. However, it was observed that the 4,7-dimethyl substituents (trans) of compound 33 are different from other di-substituents in these analogues. Compounds 39 and 40 seem very identical to each other, except that the difference occurs in Y-substituent (9-Cl). Both are

*Unpublished data.

Table 7. Experimental and calculated log (1/*C*) HIV-1 RT inhibitory affinities of training TIBO compounds, based on CoMFA analysis

Compd No.	Expt. log(1/*C*)	Calc. log(1/*C*)[a]	Residual
1	7.340	7.305	0.035
2	6.790	7.197	−0.407
3	4.300	3.806	0.494
4	5.000	4.603	0.397
5	5.000	5.089	−0.089
6	4.640	5.078	−0.438
7	4.490	4.661	−0.171
8	6.170	6.182	−0.012
9	5.660	5.576	0.084
10	4.130	4.300	−0.170
11	4.900	4.945	−0.045
12	3.740	4.073	−0.333
13	4.320	4.190	0.130
14	4.080	3.754	0.326
15	4.920	5.772	−0.852
16	6.840	7.065	−0.225
17	6.790	6.477	0.313
18	5.610	5.349	0.261
19	7.110	7.130	−0.020
20	7.920	8.266	−0.346
21	7.640	7.000	0.640
22	4.250	4.115	0.135
23	5.650	5.613	0.037
24	4.870	5.519	−0.649
25	4.840	5.061	−0.221
26	4.300	5.039	−0.739
27	5.000	4.724	0.276
28	7.380	6.495	0.885
29	5.940	6.449	−0.509
30	6.640	6.070	0.570
31	6.320	6.070	0.250
32	5.300	5.689	−0.389
33	4.590	5.595	−1.005
34	5.420	5.703	−0.283
35	5.700	6.321	−0.621
36	8.300	8.296	0.004
37	6.740	6.463	0.277
38	7.370	7.630	−0.260
39	7.470	6.469	1.001
40	7.220	6.520	0.700
41	4.220	3.791	0.429
42	5.780	5.410	0.370
43	4.460	4.762	−0.302
44	7.010	6.997	0.013
45	5.480	5.977	−0.497
46	7.580	6.605	0.975

[a] Calculated by CoMFA model 2.

Table 8. Structure and predicted log (1/*C*) HIV-1 RT inhibitory affinities of the tested TIBO compounds

Cpds	*R*	*X*	*Y*	*Z*	Expt. log(1/*C*)	Calc. log(1/*C*)[a]	Residual
[b]T1	H	O	H	DMA	4.900	5.663	−0.763
[b]T2	H	O	H	2-MA	4.330	4.405	−0.075
[b]T3	H	O	H	CH2CH2CH3	4.050	4.037	0.013
[b]T4	H	O	H	CH2C(C2H5)=CH2	4.430	4.423	0.003
[c]T5	5-CH3 (S)	S	H	DMA	7.355	7.248	0.107
[b]T6	5-CH3 (S)	O	H	CH2CH=CH2	4.154	4.796	−0.642
[b]T7	5-CH3 (S)	O	H	CH2CH2CH2CH3	3.999	4.029	−0.030
[c]T8	5-CH3 (S)	S	8-F	DMA	8.235	7.613	0.622
[c]T9	5-CH3 (S)	O	8-Br	DMA	7.324	7.501	−0.177
[c]T10	5-CH3 (S)	S	8-Br	DMA	8.521	8.273	0.248
[c]T11	5-CH3 (S)	S	8-CH3	DMA	7.865	7.338	0.527
[c]T12	5-CH3 (S)	S	8-O-CH3	DMA	7.468	7.256	0.212
[c]T13	5-CH3 (S)	S	9,10-di Cl	DMA	7.592	6.938	0.654
[c]T14	5-CH3(S)	O	8-CN	DMA	5.940	6.051	−0.111
[c]T15	5-CH3(S)	S	8-CN	DMA	7.250	6.630	0.620
[c]T16	CH3(S)	O	8-CH3	DMA	6.000	6.730	−0.730
[c]T17	5-CH3(S)	S	10-OCH3	DMA	5.330	6.352	−1.022
[c]T18	5-CH3(S)	O	10-OCH3	DMA	5.180	5.906	−0.726
[c]T19	5-CH3(S)	S	10-Br	DMA	5.970	6.850	−0.880
[c]T20	5-CH3(S)	S	8-CHO	DMA	6.730	6.548	0.182
[c]T22	15-CH3(S)	O	8-I	DMA	7.060	7.020	0.040
[c]T22	5-CH3(S)	S	8-I	DMA	7.320	7.120	0.200
[c]T23	5-CH3(S)	O	8-C=CH	DMA	6.360	6.293	0.067
[c]T24	5-CH3(S)	S	8-C=CH	DMA	7.530	6.489	1.041

[a]Calculated by CoMFA model 8.
[b]Reference 30.
[c]Reference 28.

similar to compound 38. The similarity of compound 46 is also involved by the 5-CH3(S) substitution as shown in compound 45. Finally, CoMFA cannot distinguish the structural similarity of compounds 15 and 14. Steric and electrostatic contributions of model 8 are 81.8% and 18.2%, respectively, with $r^2_{cv} = 0.771$, $S_{PRESS} = 0.612$, noc = 3. Other statistical results obtained are that the conventional r^2 is 0.941, standard error of estimate is 0.312, F is 195.028 and probability (P) of obtaining this value of F if r^2 were actually zero (prob. of $r^2 = 0$) is lower than 0.001. The non-cross-validated analysis of model 8 is plotted in Figure 4.

Prediction for compounds in the test set

The obtained CoMFA model (model 8) was used to predict the inhibitory activity of compounds in the test set. The observed and predicted inhibitory activities of 24 compounds (T1-T24) are listed in Table 8. The residual values indicated that our CoMFA model can predict the activity of TIBO derivatives not included in the training set. The model accurately predicted the activities for compounds T2, T3, T4, T7, T21 and T23 and generally predicted for T5, T6, T8-T15, T20 and T22. Comparison of predicted and experimental activities of the test set is plotted in Figure 5.

Steric and electrostatic contributions

The CoMFA steric and electrostatic fields for all 46 TIBO compounds are summarized as contour maps in Figures 6 and 7. Steric and electrostatic interactions between enzyme-inhibitor are also considered, therefore, amino acid residues surrounding the TIBO compound in the complex structure were merged into both figures. In Figure 6, the green and yellow contour maps represent regions of space whose occupancy by ligand, respectively, increases or decreases the receptor binding affinity. There are favorable steric regions corresponding to the location around the N6 side chain in the diazepine ring. Furthermore, there is a slightly

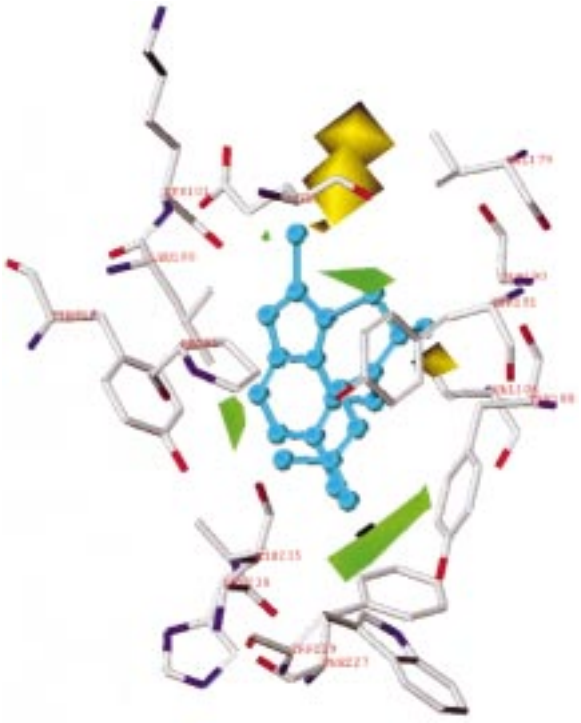

Figure 6. CoMFA steric STDEV*COEFF contour plot from the analysis based on the 3D-QSAR model 8 with no cross-validation. Sterically favored areas are represented by green regions. Sterically unfavored areas are represented by yellow regions. Compound 36 is displayed inside the fields as a ball and stick representation.

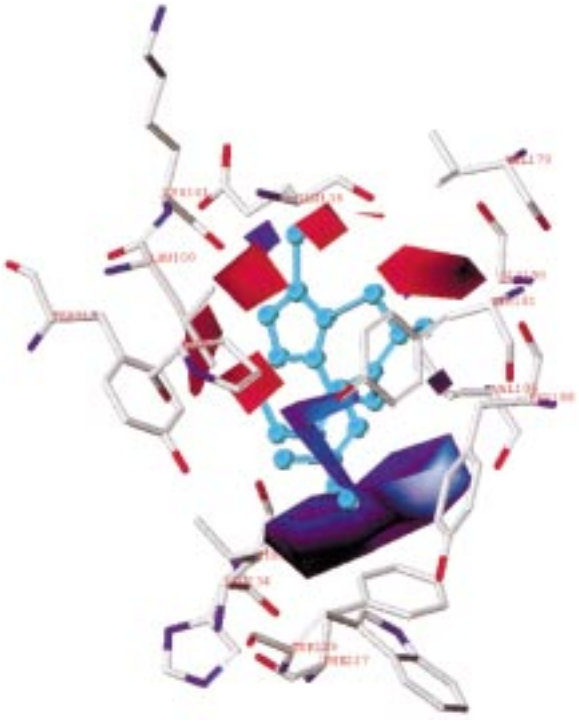

Figure 7. CoMFA electrostatic STDEV*COEFF contour plot from the analysis based on the 3D-QSAR model 8 with no cross-validation. Negative charge favored areas are represented by red regions. Negative charge unfavored areas are represented by blue regions. Compound 36 is displayed inside the fields as a ball and stick representation.

green contour located at the substituent attached to the C8-position. It is indicated that an additional bulky atom at this site would increase the activity but the dimension of this substituent should not be too large. This suggestion agrees with the trend observed experimentally that the dimethylallyl group is the optimal group of N6-substitution for inhibitory activity [30]. One of the positive steric regions, the green contour region near the position of Tyr181, one of the important amino acid residues in the binding site, may suggest that there are steric interactions between the aromatic ring of Tyr181 and N6 side chain, i.e. dimethylallyl group, and CH_3 attached to the C5 position. That means losing favorable interactions between the aromatic ring of Tyr181 and the TIBO compound may decrease the affinity of this inhibitor. This is also in agreement with the experimental results [9]. The mutation of Tyr181 $\rightarrow$ Cys apparently eliminates favorable contacts of the aromatic ring of the tyrosine and the bound inhibitor, reducing the NNRTI binding. These results can reveal the importance of the steric feature of molecules contributing to affinity through contour maps. It should be mentioned here that the contribution of the C5 parameter can not be estimated in QSAR analysis. Therefore, CoMFA results provide an opportunity to serve with this factor.

The electrostatic contribution contour map is depicted in Figure 7. The positive electrostatic contours are shown in blue and the negative contours are shown in red. A blue electrostatic contour region located at the substituent attached to the C2 position and the larger blue regions located at C8 and C9 substitute positions produce the favorable positive charges. These results complement the obtained QSAR model, particularly the relationship between the net atomic charge and the nature of the substituent on the atom. The low positive charges of the C2 atom in all compounds and a negative coefficient of this variable in Equation 5 indicate that the lower the positive charge, the higher the inhibitory activity. Red electrostatic contour regions, close to the benzene ring and close to the imidazole moiety, suggest that high negative charges in these areas enhance affinity.

Crystal structures of the TIBO compound complexed with wild-type and mutant HIV-1 RT [9] strongly suggest that there are a number of interactions of amino acid residues in the area surrounding the bound inhibitor with some parts of TIBO such as the dimethylallyl group attached to the N6 position, Cl substitution at C8 and C9 positions of the benzene ring and the methyl group resident at the C5-position of the diazepine ring. These results show close agreement with our QSAR and CoMFA models, which suggest the active center and structural requirements for an HIV-1 RT inhibitor required in this class of TIBO derivatives. Moreover, the bulky substituents of the dimethylallyl group substituted at the N6-position could also indicate favorable steric interaction of these structural elements with residues in the binding pocket of HIV-1 RT, mostly hydrophobic. One of the important binding sites of the TIBO inhibitor which was shown by the CoMFA model (Figure 6) is the steric interaction of a dimethylallyl side chain positioned at N6 with Tyr 181 and Tyr 188 residues in the binding pocket of RT. Numerous studies have indicated that Tyr181Cys HIV-1 RT mutation is frequently found in the presence of NNRTIs, this presence leads to a resistance to the NNRTIs due to the loss of hydrophobic interaction of the TIBO inhibitor. Furthermore, a Cl atom substituted in the C8- and C9-positions of the benzene ring, resulting in a higher hydrophobicity of the molecule [31], may improve interactions between the hydrophobic binding pocket of HIV-1 RT and the inhibitor which result in a favorable inhibitor. This influences the catalytic efficiency on HIV-1 RT.

Conclusions

The analysis of a set of TIBO derivatives, QSAR and CoMFA studies successfully explain the structure-activity relationships. The reasonable models, based on both statistical significance and predictive ability, derived from two methodologies reinforce each other. Furthermore, the obtained CoMFA results can reveal the importance of steric features of the inhibitor contributing to affinity through contour maps. It also indicates that losing favorable interactions between the aromatic ring of Tyr181 and the TIBO compound may decrease the affinity of this inhibitor. Consequently, these results can be useful in identifying the structural requirements of TIBO inhibitors and helpful for better understanding the anti HIV-1 RT inhibition. Eventually, these provide a beneficial basis to design new and more potent inhibitors of HIV-1 RT.

Acknowledgements

The authors would like to express their appreciation to Dr. G. Ecker for computer technical assistance, Dr. A. Beyer for providing some data on the HIV-1 RT and inhibitor complex structures, Luckhana Lawtrakul for helpful assistance and K. Peacock for reading of the manuscript. This work was supported by grants from the National Research Council of Thailand and Fonds zur Foerderung der Wissenschaftlichen Forschung (P12257) under the Austria-Thailand Cooperative Science Program (NRCT-FWF). S.H. is grateful to the Thailand Research Fund for a research fellowship (RSA4080022). The high performance computing center of the National Electronics and Computer Technology (NECTEC) is also gratefully acknowledged for providing SGI and software facilities.

References

1. Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaechers, A., Gelder, J.V., Woestenborghs, R., Heykants, J., Schellekens, K., Janessen, M.A.C., De Clercq, E. and Janssen, P.A.J., Nature, 343 (1990) 470.
2. Frank, K.B., Noll, G.J., Connell, E.V. and Sim, I.S., J. Biol. Chem., 266 (1991) 14232.
3. Althaus, I.W., Chou, J.J., Gonzales, A.J., Deibel, M.R., Chou, K.-C., Kezdy, F.J., Romero, D.L., Aristoff, P.A., Tarpley, W.G. and Reusser, F., J. Biol. Chem., 268 (1993) 6119.
4. Spence, R.A., Kati, W.M., Anderson, K.S. and Johnson, K.A., Science, 267 (1995) 988.
5. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A., Science, 256 (1992) 1783.
6. Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A.D.J., Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L, Janssen, P.A.J., Smith, R.H.J., Kroeger Koepke, R., Michejda, C.J., Hughes, S.H. and Arnold, E., Structure, 3 (1995) 365.
7. Ren, J.S., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D. and Stuart, D., Structure, 3 (1995) 915.
8. Richman, D., Shih, C.-K., Lowy, I., Rose, J., Prodanovich, P., Goff, S. and Griffin, J., Proc. Natl. Acad. Sci. USA, 88 (1991) 11241.
9. Das, K., Ding, J., Hsiou, Y., Clark Jr., A.D., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, R.A.J., Boyer, P.L., Clark, P., Smith Jr., R.H., Smith, M.B.K., Michejda, C.J., Hughes, S.H. and Arnold, E., J. Mol. Biol., 264 (1996) 1085.
10. Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R.G., Boyer, P.L., Hughes, S.H., Pauwels, R., Andries, K., Janssen, P.A.J. and Arnold, E., J. Mol. Biol., 243 (1994) 369.
11. Hannongbua, S., Lawtrakul, L. and Limtrakul, J., J. Comput.-Aided Mol. Design, 10 (1996) 145.

12. Hannongbua, S., Lawtrakul, L., Sotriffer, C.A. and Rode, B.M., Quant. Struct.-Act. Relat., 15 (1996) 389.
13. Hansch, C. and Fujita, T., J. Am. Chem. Soc., 86 (1964) 1616.
14. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
15. Breslin, H.J., Kukla, M.J., Ludovici, D.W., Mohrbacher, R., Ho, W., Miranda, M., Rodgers, J.D., Hitchens, T.K., Leo, G., Gauthier, D.A., Ho, C.Y., Scott, M.K., De Clercq, E., Pauwels, R., Andries, K., Janssen, M.A.C. and Janssen, P.A.J., J. Med. Chem., 38 (1995) 771.
16. Tripos Associates Inc., St. Louis, MO.
17. Dewar, M.J., Zoebisch, E.G., Healy, E.F. and Stewart, J.P., J. Am. Chem. Soc., 107 (1985) 3902 .
18. GAUSSIAN 94, Pittsburgh, PA, 1995.
19. ChemPlus 1.0, Hypercube Inc., Waterloo, ON, USA, 1993.
20. Miller, K.J., J. Am. Chem. Soc., 112 (1990) 8533.
21. Norusis, M.J., SPSS for Windows Release 6.0., SPSS, Inc., Chicago, 1993.
22. Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN.
23. SYBYL Molecular Modelling Software, version 6.4, Tripos Associates, Inc., St. Louis, MO. 1996.
24. Wold, S., Johansson, E. and Cocchi, M., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 523-549.
25. SYBYL Molecular Modelling Software, version 6.3, SYBYL Ligand Base Design, Tripos Associates, Inc., St. Louis, MO, 1996, p. 229.
26. Topliss, J.G. and Costello, R.J., J. Med. Chem., 15 (1972) 1066.
27. Wold, S., Quant. Struct.-Act. Relat., 10 (1991) 191 .
28. Ho, W., Kukla, M.J., Breslin, H.J., Lodevici, D.W., Grows, P.P., Diamond, C.J., Miranda, M., Rodgers, D.W., Ho, C.Y., De Clercq, E., Pauwels, R., Andries, K., Janssen, M.A.C. and Janssen, P.A.J., J. Med. Chem., 38 (1995) 794.
29. Kukla, M.J., Breslin, H.J., Diamond, C.J., Grods, P.P., Ho, C.Y., Miranda, M., Roders, J.D., Sherril, R.G., De Clercq, E. and Janssen, P.A.J., J. Med. Chem., 34 (1991) 3187.
30. Kukla, M.J., Breslin, H.J., Pauwels, R., Fedde, C.L., Scott, M.K., Miranda, M., Sherril, R.G., Raeymaeker, A., Gelder, J.V., Andries, K., Janssen, M.A.C., De Clercq, E. and Janssen, P.A.J., J. Med. Chem., 34 (1991) 746.
31. Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley Interscience, New York, NY, 1979.