Molecular docking and quantitative structure activity relationship (QSAR) analyses of indolylarylsulfones as HIV-1 non-nucleoside reverse transcriptase inhibitors. Med Chem Res (in...

ARTICLE *in* MEDICINAL CHEMISTRY RESEARCH · JANUARY 2013

Impact Factor: 1.4 · DOI: 10.1007/s00044-013-0647-8.

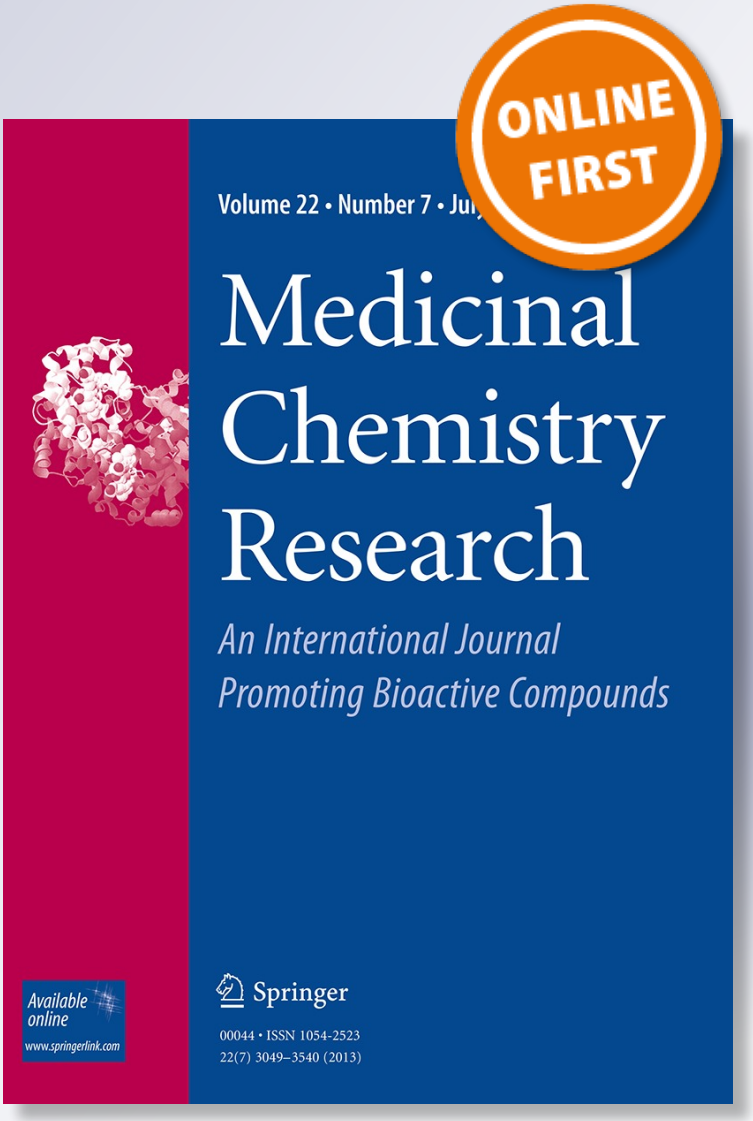
Molecular docking and quantitative structure–activity relationship (QSAR) analyses of indolylarylsulfones as HIV-1 non-nucleoside reverse transcriptase inhibitors

Vijay H. Masand, Devidas T. Mahajan, Taibi Ben Hadda, Rahul D. Jawarkar, Hemant Chavan, B. P. Bandgar & Harsh Chauhan

Medicinal Chemistry Research

ISSN 1054-2523

Med Chem Res
DOI 10.1007/s00044-013-0647-8



Springer

Med Chem Res
DOI 10.1007/s00044-013-0647-8

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

Molecular docking and quantitative structure–activity relationship (QSAR) analyses of indolylarylsulfones as HIV-1 non-nucleoside reverse transcriptase inhibitors

Vijay H. Masand · Devidas T. Mahajan · Taibi Ben Hadda · Rahul D. Jawarkar · Hemant Chavan · B. P. Bandgar · Harsh Chauhan

Received: 16 January 2013 / Accepted: 22 May 2013

Abstract Indolylarylsulfones (IASs) have received considerable interest during the last decades due to high potency against HIV-1 as non-nucleoside reverse transcriptase inhibitors. In present work, quantitative structure–activity relationship (QSAR) and molecular docking analyses were performed to model the anti-HIV-1 activity of 36 IASs. 2D and 3D-descriptors, genetic algorithm, internal and external validations were used to develop statistically robust four-parametric QSAR models. The best QSAR model is with $R^2_{tr} = 0.8608$. The QSAR analysis reveals that the activity of IASs depends on the presence of electronegative and heavy atoms at the internal atmosphere of the compounds. The docking analysis reveals that lipophilic and H-bonding interactions are the prominent interactions among IASs and the receptor. The QSAR analysis proved to be a useful tool in the prediction of anti-HIV-1 activity of congeneric compounds and some important insights were also found that will be useful to guide for the synthesis of new IASs with improved activity.

Keywords Molecular docking · QSAR · Indolylarylsulfones · HIV-1

Abbreviations

QSAR	Quantitative structure–activity relationship
NNRTIs	Non-nucleoside reverse transcriptase inhibitors
IASs	Indolylarylsulfones
WHO	World Health Organization
AIDS	Acquired Immunodeficiency Syndrome
HAART	Highly active antiretroviral therapy
RT	Reverse transcriptase
GA	Genetic algorithm

Introduction

Human Immunodeficiency Virus (HIV), the etiological agent of Acquired Immunodeficiency Syndrome (AIDS), is a major health problem in many countries. World Health Organization (WHO) and many other agencies are actively engaged in prevention and treatment of this dreadful disease (WHO Report, 2011). New drugs and therapies are being developed to diagnose, prevent, control, and cure this lethal disease. The life cycle of HIV has been remarkably analyzed, which has led to identification of many of its steps as attractive targets for drug designing. The highly active antiretroviral therapy (HAART), in which three (recommended) or four different antiretroviral drugs are used, for HIV inhibition is chiefly based on the inhibition of three key viral enzymes: HIV-1 reverse transcriptase

V. H. Masand (✉) · D. T. Mahajan
Department of Chemistry, Vidya Bharati College, Camp, Amravati, Maharashtra, India
e-mail: vijaymasand@gmail.com; vijaymasand@rediffmail.com

T. B. Hadda
Laboratoire Chimie des Matériaux, Université Mohammed Premier, 60000 Oujda, Morocco

R. D. Jawarkar
Department of Pharmaceutical Chemistry, P. Wadhwani College of Pharmacy, Dhamangaon Rly. Road, Yavatmal, Maharashtra, India

H. Chavan
Department of Chemistry, Solapur University, Solapur, Maharashtra, India

B. P. Bandgar · H. Chauhan
Creighton University, Greater Boston Area, Omaha, USA

Published online: 05 June 2013

(RT), HIV-1 protease, and HIV-1 integrase. Among these proteins that are crucial in life cycle of HIV, the inhibition of RT, which does not exist in Human and play an important role in the viral replication, is considered as one of the most attractive target in the anti-HIV chemotherapy. The emergence of resistance against many drugs used in HAART is a cause of serious concern. Thereby, development of new drugs or modification of existing drugs is essential to restrain HIV-1. These important goals can be achieved using modern methods of drug designing viz. QSAR, Molecular docking, pharmacophore modeling, etc. (Mahajan *et al.*, 2012, 2013; Masand *et al.*, 2010, 2012, 2013).

Among these techniques, QSAR and molecular docking are thriving techniques that have gained considerable attention because of good success. The central theme and application of QSAR is to establish the correlation between structure and biological response, whereas molecular docking is performed to determine the structural features that are important for interaction with a receptor (Doweyko, 2008; Mahajan *et al.*, 2012, 2013; Masand *et al.*, 2010, 2012, 2013). In present study, we have performed QSAR and docking analysis to determine the pharmacophoric features and conformation that steers the anti-HIV-1 activity of IASs.

Dataset and descriptor calculation

The dataset selected (La Regina *et al.*, 2011) for present study includes 36 IASs possessing a variety of substituents at various positions (Fig. 1; Table 1). The HIV-1 RT inhibitory activity reported as IC_{50} (nM) was converted to $-\log_{10}IC_{50} = pIC_{50}$ (M). The structures were drawn and optimized using default settings in ACD ChemSketch 12. e-Dragon was used to calculate variety of 2D and 3D descriptors.

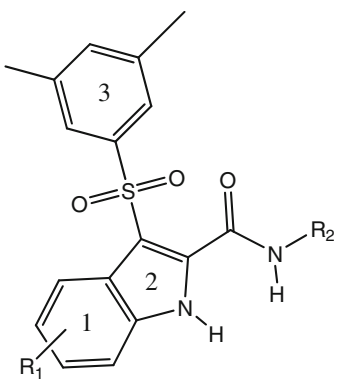


Fig. 1 IASs used in present study

QSAR model development

Genetic algorithm (GA) module available in QSARINS (Chirico and Gramatica, 2011, 2012; Gramatica, 2013) was used to select optimum number and set of descriptors. In present analysis, two QSAR equations were developed for complete dataset ($N = 36$) and by dividing it into training set ($N_{tr} = 29$) and test set ($N_{ex} = 7$) randomly, for better external validation (Golbraikh and Tropsha, 2002; Gramatica *et al.*, 2007; Gramatica and Papa, 2007; Mahajan *et al.*, 2012, 2013; Martin *et al.*, 2012; Masand *et al.*, 2010, 2012, 2013; Sushko *et al.*, 2010; Tropsha 2010).

QSAR models

The GA-MLR QSAR models along with their statistical characteristics and interpretation are as following:

$$pIC_{50} = 10.7963 - 0.0549RDF105m - 17.9976G2e + 0.1892H - 048 + 0.1604F07[C-S] \quad \text{(Model 1)}$$

$N_{tr} = 36$, $Q^2_{loo} = 0.7779$, $R^2_{tr} = 0.8608$, $R^2_{adj} = 0.8428$, $K_{xx} = 0.3179$, $\Delta K = 0.1425$, $RMSE_{tr} = 0.1595$, $RMSE_{cv} = 0.2015$, $Sy = 0.1719$, $F = 47.9201$, $CCC_{tr} = 0.9252$, $CCC_{cv} = 0.8795$, $MAE_{tr} = 0.1227$, $MAE_{cv} = 0.1488$, $RSS_{tr} = 0.9160$, $PRESS_{cv} = 1.4612$, $R^2_{LMO} = 0.8695$, $Q^2_{LMO} = 0.7999$.

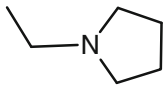
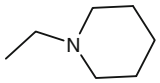
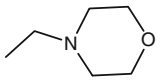
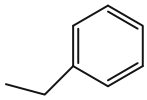
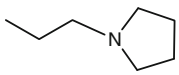
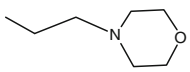
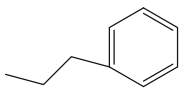
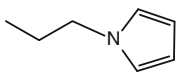
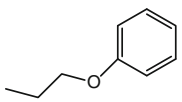
$$pIC_{50} = 11.1047 - 0.0606RDF105m - 19.8127G2e + 0.2046H - 048 + 0.1602F07[C-S] \quad \text{(Model 1a)}$$

$N_{tr} = 29$, $N_{ex} = 7$, $Q^2_{loo} = 0.7465$, $R^2_{tr} = 0.8674$, $R^2_{adj} = 0.8452$, $K_{xx} = 0.2749$, $\Delta K = 0.1439$, $RMSE_{tr} = 0.1507$, $RMSE_{cv} = 0.2083$, $RMSE_{ex} = 0.2058$, $Sy = 0.1656$, $F = 39.2325$, $Q^2{-}F^1 = 0.8223$, $Q^2{-}F^2 = 0.7872$, $Q^2{-}F^3 = 0.7525$, $CCC_{tr} = 0.9290$, $CCC_{cv} = 0.8639$, $CCC_{ex} = 0.9036$, r^2m av $= 0.7928$, r^2m de $= 0.1061$, $MAE_{tr} = 0.1196$, $MAE_{cv} = 0.1558$, $MAE_{ex} = 0.1515$, $RSS_{tr} = 0.6585$, $PRESS_{cv} = 1.2582$, $PRESS_{ex} = 0.2966$, $R^2_{LMO} = 0.8785$, $Q^2_{LMO} = 0.7569$ (Table 2).

$$pIC_{50} = 21.5026 - 3.6217BEHm7 - 0.0319RDF045m - 0.0461RDF025e + 0.1815F07[N-O] \quad \text{(Model 2)}$$

$N_{tr} = 36$, $Q^2_{loo} = 0.7810$, $R^2_{tr} = 0.8502$, $R^2_{adj} = 0.8309$, $K_{xx} = 0.3053$, $\Delta K = 0.1369$, $RMSE_{tr} = 0.1654$, $RMSE_{cv} = 0.2000$, $Sy = 0.1783$, $F = 44.0019$, $CCC_{cv} = 0.8820$, $MAE_{tr} = 0.1307$, $MAE_{cv} = 0.1555$, $RSS_{tr} = 0.9854$, $PRESS_{cv} = 1.4407$, $R^2_{LMO} = 0.8576$, $Q^2_{LMO} = 0.7831$.

Table 1 IASs used in present work along with their pIC_{50} values

S. no.	R^1	R^2	IC_{50} (nM)	pIC_{50} (M)
1	5-Cl		3.3	8.481
2	5-Br		1.3	8.886
3	5-NO_2		2.5	8.602
4	5-Cl, 4-F		3.9	8.409
5	5-Cl		1.3	8.886
6	5-Br		3.7	8.432
7	5-NO_2		3.1	8.509
8	5-Cl, 4-F		3.9	8.409
9	5-Cl		1.9	8.721
10	5-Br		3.4	8.469
11	5-NO_2		5.8	8.237
12	5-Cl, 4-F		2.5	8.602
13	5-Cl		5.7	8.244
14	5-Br		5.7	8.244
15	5-NO_2		6.2	8.208
16	5-Cl, 4-F		5.8	8.237
17	5-Cl		17	7.77
18	5-Br		16	7.796
19	5-NO_2		16	7.796
20	5-Cl, 4-F		28	7.553
21	5-Cl		8.8	8.056
22	5-Br		170	6.77
23	5-NO_2		8	8.097
24	5-Cl, 4-F		13	7.886
25	5-Cl		5.7	8.244
26	5-Br		14	7.854
27	5-NO_2		6.8	8.167
28	5-Cl, 4-F		14	7.854
29	5-Cl		6.5	8.187
30	5-Br		6.4	8.194
31	5-NO_2		6.9	8.161
32	5-Cl, 4-F		5.5	8.26
33	5-Cl		26	7.585
34	5-Br		29	7.538
35	5-NO_2		29	7.538
36	5-Cl, 4-F		19	7.721

$$pIC_{50} = 21.5026 - 3.6217BEHm7 - 0.0319RDF045m - 0.0461RDF025e + 0.1815F07[N-O] \quad \text{(Model 2a)}$$

$N_{tr} = 29$, $N_{ex} = 7$, $Q^2_{loo} = 0.7531$, $R^2_{tr} = 0.8535$, $R^2_{adj} = 0.8291$, $K_{xx} = 0.2540$, $\Delta K = 0.1597$, $RMSE_{tr} = 0.1680$, $RMSE_{cv} = 0.2181$, $RMSE_{ex} = 0.1675$, Sy = 0.1847, $F = 34.9549$, $Q^2{-}F^1 = 0.8097$, $Q^2{-}F^2 = 0.7633$, $Q^2{-}F^3 = 0.8543$, $CCC_{tr} = 0.9210$, $CCC_{cv} = 0.8676$, $CCC_{ex} = 0.8874$, r^2m av = 0.7674, r^2m de = 0.0137, $MAE_{tr} = 0.1273$, $MAE_{cv} = 0.1599$, $MAE_{ex} = 0.1482$, $RSS_{tr} = 0.8185$, $PRESS_{cv} = 1.3796$, $PRESS_{ex} = 0.1965$, $R^2_{LMO} = 0.8641$, $Q^2_{LMO} = 0.7664$ (Table 3).

The symbols have their usual meaning (Chirico and Gramatica, 2011, 2012; Gramatica 2013; Gramatica *et al.*, 2007; Kim *et al.*, 2007; Kovarich *et al.*, 2012; Liu *et al.*, 2008; Mahajan *et al.*, 2013; Masand *et al.*, 2010, 2012, 2013). The high values of R, R^2, R^2_{adj}, R^2_{cv}, and F with low value of Sy, $RMSE_{train}$, and PRESS confirm statistical robustness of the equation. The close values of R^2, R^2_{cv}, R^2_{L10O}, and R^2_{L5O} reveal that the equation is statistically

Table 2 Experimental and predicted pIC_{50} by Model 1

ID	Status	Exp. pIC_{50}	Pred. fitting pIC_{50}	Pred.fit. res.	Pred. LOO	Pred. LOO res.	HAT i/i
1	Training	8.4810	8.8284	0.3474	8.8830	0.4020	0.1358
2	Training	8.8860	8.6661	−0.2199	8.6452	−0.2408	0.0870
3	Training	8.6020	8.6317	0.0297	8.6352	0.0332	0.1064
4	Training	8.4090	8.3580	−0.0510	8.3508	−0.0582	0.1241
5	Training	8.8860	8.7696	−0.1164	8.7527	−0.1333	0.1264
6	Training	8.4320	8.4582	0.0262	8.4635	0.0315	0.1661
7	Training	8.5090	8.5530	0.0440	8.5567	0.0477	0.0774
8	Training	8.4090	8.6580	0.2490	8.6795	0.2705	0.0793
9	Training	8.7210	8.6574	−0.0636	8.6517	−0.0693	0.0815
10	Training	8.4690	8.4930	0.0240	8.4961	0.0271	0.1141
11	Training	8.2370	8.3996	0.1626	8.4204	0.1834	0.1137
12	Training	8.6020	8.5546	−0.0474	8.5509	−0.0511	0.0735
13	Training	8.2440	8.1285	−0.1155	8.1091	−0.1349	0.1435
14	Training	8.2440	8.2161	−0.0279	8.2120	−0.0320	0.1304
15	Training	8.2080	8.0715	−0.1365	8.0417	−0.1663	0.1792
16	Training	8.2370	8.2925	0.0555	8.3019	0.0649	0.1449
17	Training	7.7700	7.7414	−0.0286	7.7382	−0.0318	0.0998
18	Training	7.7960	7.8741	0.0781	7.8832	0.0872	0.1045
19	Training	7.7960	7.8371	0.0411	7.8420	0.0460	0.1068
20	Training	7.5530	7.6640	0.1110	7.6826	0.1296	0.1440
21	Training	8.0560	7.9515	−0.1045	7.9364	−0.1196	0.1265
22	Training	6.7700	7.1549	0.3849	7.4077	0.6377	0.3964
23	Training	8.0970	7.7646	−0.3324	7.7409	−0.3561	0.0667
24	Training	7.8860	7.9092	0.0232	7.9121	0.0261	0.1083
25	Training	8.2440	8.1489	−0.0951	8.1383	−0.1057	0.1010
26	Training	7.8540	7.9413	0.0873	7.9488	0.0948	0.0789
27	Training	8.1670	8.0299	−0.1371	8.0201	−0.1469	0.0667
28	Training	7.8540	7.6784	−0.1756	7.5425	−0.3115	0.4365
29	Training	8.1870	7.9708	−0.2162	7.9255	−0.2615	0.1733
30	Training	8.1940	7.9855	−0.2085	7.9430	−0.2510	0.1695
31	Training	8.1610	8.1820	0.0210	8.1860	0.0250	0.1602
32	Training	8.2600	8.2790	0.0190	8.2838	0.0238	0.2004
33	Training	7.5850	7.8934	0.3084	7.9385	0.3535	0.1275
34	Training	7.5380	7.4051	−0.1329	7.3454	−0.1926	0.3098
35	Training	7.5380	7.7250	0.1870	7.7393	0.2013	0.0710
36	Training	7.7210	7.7304	0.0094	7.7311	0.0101	0.0691

stable apropos to inclusion–exclusion of molecules in dataset. In addition, the value of CCCex is higher than the threshold value (>0.85) that indicates high external predictivity of the models (Chirico and Gramatica, 2011, 2012; Gramatica 2013; Gramatica *et al.*, 2007; Kim *et al.*, 2007; Liu and Gramatica, 2007; Liu *et al.*, 2008). The plot between experimental and predicted pIC_{50} (depicted in Fig. 2) reveals a good relation and predictive ability of the equation.

Results and discussions

Validation and interpretation of QSAR models

Appropriate validation and interpretation of QSAR models are very crucial to prove that the models are very useful for steering the future lead optimization. For the QSAR models developed in present study, apparently the anti-HIV activity varies with 2D and 3D descriptors viz. *RDF*105*m* (Radial

Table 3 Experimental and predicted pIC_{50} by model 2

ID	Status	Exp. pIC_{50}	Pred. fitting pIC_{50}	Pred. fit. res.	Pred. LOO	Pred. LOO res.	HAT i/i
1	Training	8.4810	8.5594	0.0784	8.5655	0.0845	0.0732
2	Training	8.8860	8.5972	−0.2888	8.5748	−0.3112	0.0721
3	Training	8.6020	8.8402	0.2382	8.8967	0.2947	0.1917
4	Training	8.4090	8.4614	0.0524	8.4659	0.0569	0.0791
5	Training	8.8860	8.6337	−0.2523	8.6123	−0.2737	0.0780
6	Training	8.4320	8.3630	−0.0690	8.3567	−0.0753	0.0838
7	Training	8.5090	8.3950	−0.1140	8.3680	−0.1410	0.1916
8	Training	8.4090	8.4314	0.0224	8.4340	0.0250	0.1034
9	Training	8.7210	8.4457	−0.2753	8.4208	−0.3002	0.0830
10	Training	8.4690	8.4908	0.0218	8.4925	0.0235	0.0755
11	Training	8.2370	8.4536	0.2166	8.4977	0.2607	0.1691
12	Training	8.6020	8.5341	−0.0679	8.5285	−0.0735	0.0763
13	Training	8.2440	8.2826	0.0386	8.2876	0.0436	0.1136
14	Training	8.2440	8.2675	0.0235	8.2703	0.0263	0.1068
15	Training	8.2080	8.1797	−0.0283	8.1779	−0.0301	0.0598
16	Training	8.2370	8.1567	−0.0803	8.1478	−0.0892	0.1004
17	Training	7.7700	7.9690	0.1990	7.9912	0.2212	0.1002
18	Training	7.7960	7.7137	−0.0823	7.7005	−0.0955	0.1386
19	Training	7.7960	7.9751	0.1791	7.9838	0.1878	0.0458
20	Training	7.5530	7.6461	0.0931	7.6736	0.1206	0.2277
21	Training	8.0560	7.8979	−0.1581	7.8685	−0.1875	0.1564
22	Training	6.7700	7.1359	0.3659	7.2993	0.5293	0.3087
23	Training	8.0970	7.9138	−0.1832	7.8712	−0.2258	0.1886
24	Training	7.8860	7.7441	−0.1419	7.7121	−0.1739	0.1839
25	Training	8.2440	8.1320	−0.1120	8.1173	−0.1267	0.1159
26	Training	7.8540	7.8517	−0.0023	7.8512	−0.0028	0.1706
27	Training	8.1670	8.2215	0.0545	8.2287	0.0617	0.1162
28	Training	7.8540	7.7434	−0.1106	7.7207	−0.1333	0.1705
29	Training	8.1870	8.3979	0.2109	8.4341	0.2471	0.1466
30	Training	8.1940	8.1203	−0.0737	8.1040	−0.0900	0.1806
31	Training	8.1610	8.5957	0.4347	8.6679	0.5069	0.1425
32	Training	8.2600	8.1963	−0.0637	8.1803	−0.0797	0.2017
33	Training	7.5850	7.6314	0.0464	7.6387	0.0537	0.1356
34	Training	7.5380	7.3513	−0.1867	7.2575	−0.2805	0.3345
35	Training	7.5380	7.6156	0.0776	7.6345	0.0965	0.1957
36	Training	7.7210	7.6584	−0.0626	7.6528	−0.0682	0.0820

Distribution Function—10.5/weighted by atomic masses), *G2e* (2nd component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities), *H*-048 (H attached to C2(sp^3)/C1(sp^2)/C0(sp)), *F*07[*C*–*S*] (frequency of occurrence of C–S at a topological distance of seven), *BEHm*7 (highest eigen value *n*. 7 of Burden matrix/weighted by atomic masses), *RDF*045*m* (Radial Distribution Function—4.5/weighted by atomic masses), *RDF*025*e* (Radial Distribution Function—2.5/weighted by atomic Sanderson electronegativities), and *F*07[*N*–*O*] (frequency of occurrence of N–O at a topological distance of seven).

Interpretation of model 1 and 1a

The positive coefficients for *H*-048 and *F*07[*C*–*S*] reveal that they contribute positively toward anti-HIV activity whereas reverse is true for *RDF*105*m* and *G2e*. This fact is supported by the comparison of **1** (EC_{50} = 3.3 nM) with **5** (EC_{50} = 1.3 nM), **14** (EC_{50} = 5.7 nM) with **26** (EC_{50} = 14 nM), **2** (EC_{50} = 1.3 nM) with **18** (EC_{50} = 16 nM), and **10** (EC_{50} = 3.4 nM) with **22** (EC_{50} = 170 nM) as well as **13–16** (EC_{50} = 5.7–6.2 nM) with **33–36** (EC_{50} = 19–26 nM).

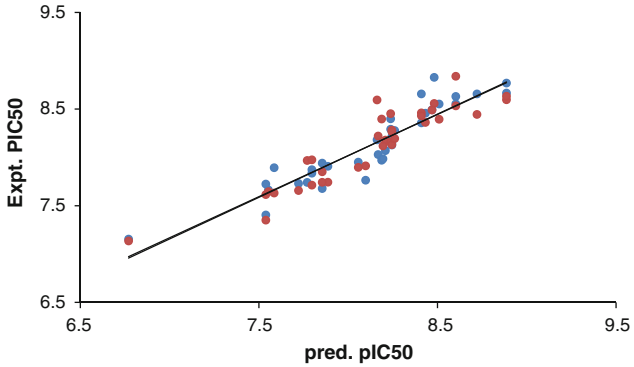


Fig. 2 Plot between experimental and predicted pIC_{50} by model 1 (*blue*) and 2 (*red*) (Color figure online)

Interpretation of model 2 and 2a

The positive coefficient of *F*07[*N–O*] shows that increase in the frequency of occurrence of N–O at topological distance of seven is beneficial for escalating the activity. This is confirmed by the high activity of compounds **1–16**, having high occurrence of N–O at topological distance of seven, than **17–36**. The negative coefficients of *BEHm*7, *RDF*045*m*, and *RDF*025*e* indicate that their values should be as low as possible. These observations are vindicated by the difference in the activity of **1** (EC_{50} = 3.3 nM) with **5** (EC_{50} = 1.3 nM), and **2** (EC_{50} = 1.3 nM) with **40** (EC_{50} = 16 nM).

The IASs possess varieties of conformations (Fig. 3). The correlation of biological activity with different descriptors viz. RDF and BEH indicates that IASs adapt a new conformation to have appropriate interactions with the receptor, as supported by the docking analysis.

The radial distribution function (RDF) descriptors are based on the distance distribution of the compounds. The RDF descriptors of a molecule of *n* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius R. In energy-optimized conformation, neither the heavy nor highly electronegative atoms are close. But, the correlation of activity with 3D-descriptor *RDF*045*m* reveals that during their interaction with the receptor, IASs adapt a specific conformation, as vindicated by the molecular docking analysis (Fig. 4), such that the heavy atoms are closer (within 4.5 Å) which results in internal repulsion, however, with augmentation in activity. Similarly, the descriptor *RDF*025*e* underlines the significance of the electrostatic environment. Figure 4 represents the conformations of molecules in docking pose and energy-optimized condition. As it can be observed, the

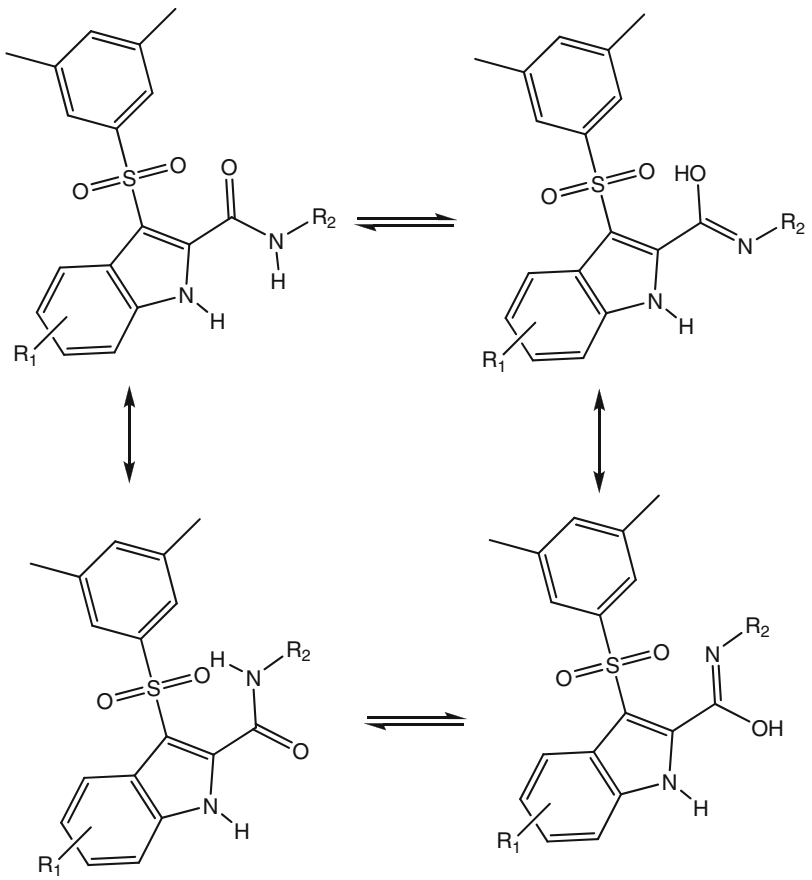


Fig. 3 Different tautomeric and conformational forms of Indole-2-carboxamides

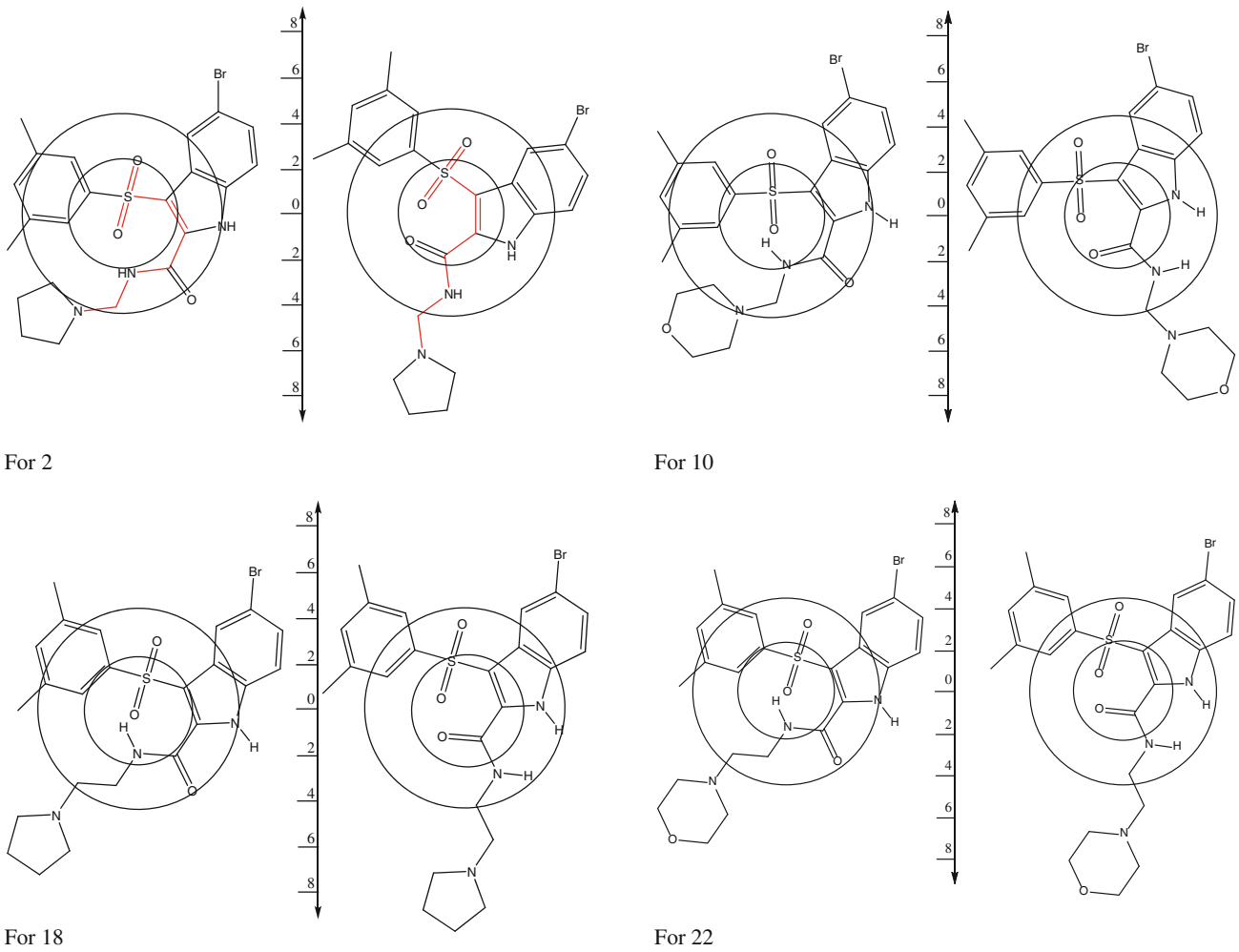


Fig. 4 Graphical representation of RDF descriptors for different conformations of **2**, **18**, **10**, and **22**, respectively, the compounds with the highest and lowest anti-HIV-1 activities

main difference is the presence of electronegative atoms like O and N in the inner regions of the compounds, specifically inside 2.5 Å.

Molecular docking

Keto form

The prominent interactions between keto form of the most active molecule **2** and receptor are arene–cation, hydrophobic and H-bonding in nature (Fig. 5). It interacts with **Lys102** and **Lys347** (2.25 Å, 14 %) due to H-bonding, with **Asp320** and **Glu344** through hydrophobic interaction, and with **Lys323** (2.35 Å, 11 %) due to arene–cation interaction (see Fig. 5). In the docking pose, the NH (ring 2) and keto group of amide are oriented on the same side forming a NH (δ–) and C=O (δ–) pharmacophore site. The rings 1 and 3 act as lipophilic and are responsible for interaction with hydrophobic region of the pocket. In addition, the water (close to **Lys102**) inside the active site acts as bridge between drug and receptor due to the H-bond formation (2.57 Å, 23 %).

Enol form

The enol form of compound **2** interacts differently with receptor than the keto form (Fig. 5). For the enol form, the major interactions are H-bond formation and hydrophobic in nature. It interacts with **Lys103** (2.3 Å, 22 %) due to H-bonding, and with **Pro176**, **Asp177**, **Ile178**, and **Val179** due to mild polar and hydrophobic interactions. Interestingly, the –OH group formed due to keto–enol tautomerization is oriented away from NH of ring 2. Surprisingly, the water molecule which plays a crucial role for the keto form does not play any role in docking of enol form.

Figure 6 depicts the summary of SAR, QSAR, and molecular docking analyses of IASs.

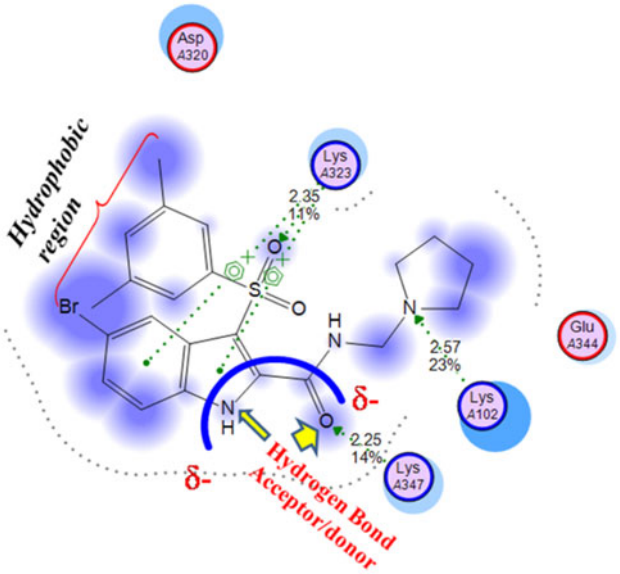


Fig. 5 Docking poses for keto and enol forms of the most active compound **1**

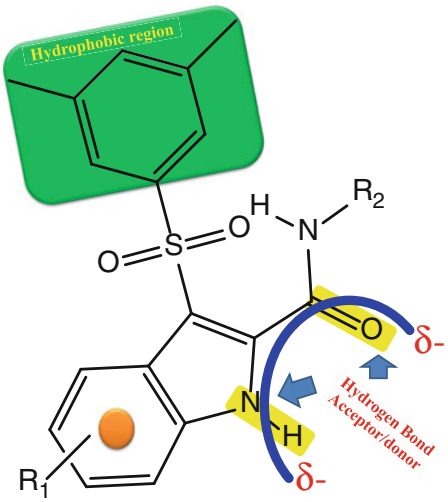


Fig. 6 Summary of SAR, QSAR, and docking analyses

From docking analysis, keto and enol form of IASs have not only different orientations inside the active site but also different amino acid residues, level, and atoms are involved in deciding the interactions. Literature survey reveals that the drugs which are capable of releasing the buried water molecule from active site of receptor possess exceptionally high binding entropy as well as gain in selectivity is also possible. Thus, the stabilization of –OH group formed due to keto–enol tautomerization by either chemical transformation or by some other method is an attractive area for significant rise in the activity and selectivity of IASs as anti-HIV activity.

Conclusions

In conclusion, the QSAR and docking analyses presented here accentuate the important structural features governing the anti-HIV activity of IASs. Moreover, they proffer supportive insinuations for the design of new IASs with improved activity profile. The models are useful to design new anti-HIV IASs with better activity, before their synthesis.

Acknowledgments We are thankful to Dr. K. N. Patil, Amravati and Dr. Ashraf Ali, Singapore for useful discussions. Sincere thanks to Dr. F. C. Raghuwanshi for encouragement and facilities.

References

Chirico N, Gramatica P (2011) Real external predictivity of QSAR models: how to evaluate it? comparison of different validation criteria and proposal of using the concordance correlation coefficient. J Chem Inf Model 51(9):2320–2335

Chirico N, Gramatica P (2012) Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J Chem Inf Model 52(8):2044–2058

Doweyko AM (2008) QSAR: dead or alive? J Comput Aided Mol Des 22(2):81–89

Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20(4):269–276

Gramatica P (2013) On the development and validation of QSAR models. Methods Mol Biol 930:499–526

Gramatica P, Papa E (2007) Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure. Environ Sci Technol 41(8):2833–2839

Gramatica P, Pilutti P, Papa E (2007) Approaches for externally validated QSAR modelling of nitrated polycyclic aromatic hydrocarbon mutagenicity. SAR QSAR Environ Res 18(1–2): 169–178

Kim JH, Gramatica P, Kim MG, Kim D, Tratnyek PG (2007) QSAR modelling of water quality indices of alkylphenol pollutants. SAR QSAR Environ Res 18(7–8):729–743

Kovarich S, Papa E, Li J, Gramatica P (2012) QSAR classification models for the screening of the endocrine-disrupting activity of perfluorinated compounds. SAR QSAR Environ Res 23(3–4): 207–220

La Regina G, Coluccia A, Brancale A, Piscitelli F, Gatti V, Maga G, Samuele A, Pannecouque C, Schols D, Balzarini J, Novellino E, Silvestri R (2011) Indolylarylsulfones as HIV-1 non-nucleoside reverse transcriptase inhibitors: new cyclic substituents at indole-2-carboxamide. J Med Chem 54(6):1587–1598

Liu H, Gramatica P (2007) QSAR study of selective ligands for the thyroid hormone receptor beta. Bioorg Med Chem 15(15): 5251–5261

Liu H, Papa E, Gramatica P (2008) Evaluation and QSAR modeling on multiple endpoints of estrogen activity based on different bioassays. Chemosphere 70(10):1889–1897

Mahajan DT, Masand VH, Patil KN, Ben Hadda T, Jawarkar RD, Thakur SD, Rastija V (2012) CoMSIA and POM analyses of anti-malarial activity of synthetic prodiginines. Bioorg Med Chem Lett 22(14):4827–4835

Mahajan DT, Masand VH, Patil KN, Hadda TB, Rastija V (2013) Integrating GUSAR and QSAR analyses for antimalarial activity of synthetic prodiginines against multi drug resistant strain. Med Chem Res 22:2284–2292

Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, Tropsha A (2012) Does rational selection of training and test sets improve the outcome of QSAR modeling? J Chem Inf Model 52(10):2570–2578

Masand VH, Jawarkar RD, Patil KN, Nazerruddin GM, Bajaj SO (2010) Correlation potential of Wiener index and molecular refractivity vis-a'-vis antimalarial activity of xanthone derivatives. Org Chem 6(1):30–38

Masand VH, Jawarkar RD, Mahajan DT, Hadda TB, Sheikh J, Patil KN (2012) QSAR and CoMFA studies of biphenyl analogs of the anti-tuberculosis drug (6S)-2-nitro-6-{[4-(trifluoromethoxy) benzyl]oxy}-6,7-dihydro-5*H*-imidazo[2,1-*b*][1,3]oxazine(PA-824). Med Chem Res 21:2624–2629

Masand VH, Mahajan DT, Patil KN, Hadda TB, Youssoufi MH, Jawarkar RD, Shibi IG (2013) Optimization of antimalarial activity of synthetic prodiginines: QSAR, GUSAR, and CoMFA analyses. Chem Biol Drug Des 81(4):527–536

Sushko I, Novotarskyi S, Korner R, Pandey AK, Cherkasov A, Li J, Gramatica P, Hansen K, Schroeter T, Muller KR, Xi L, Liu H, Yao X, Oberg T, Hormozdiari F, Dao P, Sahinalp C, Todeschini R, Polishchuk P, Artemenko A, Kuz'min V, Martin TM, Young DM, Fourches D, Muratov E, Tropsha A, Baskin I, Horvath D, Marcou G, Muller C, Varnek A, Prokopenko VV, Tetko IV (2010) Applicability domains for classification problems: benchmarking of distance to models for Ames mutagenicity set. J Chem Inf Model 50(12):2094–2111

Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inf 29:476–488

WHO Report (2011) http://www.who.int/. Accessed 23 Nov 2012