QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors

ARTICLE *in* ARABIAN JOURNAL OF CHEMISTRY · JANUARY 2013

Impact Factor: 3.73 · DOI: 10.1016/j.arabjc.2013.01.015

2 AUTHORS:

Vijay Kumar Srivastav
Shri Govindram Seksaria Institute of Techn…
8 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Meena Tiwari
Shri Govindram Seksaria Institute of Techn…
10 PUBLICATIONS **21** CITATIONS
SEE PROFILE

King Saud University

Arabian Journal of Chemistry

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors

V.K. Srivastav [1], M. Tiwari *

Computer Aided Drug Design Lab, Department of Pharmacy, Shri Govindram Seksaria Institute of Technology and Science, 23 Park Road, Indore, MP 452003, India

Received 16 July 2012; accepted 15 January 2013

KEYWORDS

Human immunodeficiency virus;
Coumarin;
Integrase inhibition;
Quantitative structure–activity relationship;
Docking

Abstract Human immunodeficiency virus integrase (HIV-1IN) is an emerging and potential drug target for anti-HIV therapy. It is an enzyme essential for 3′ processing and integration step in the life cycle of HIV. In the present study a series of coumarin derivatives (containing 26 compounds) as HIV-1IN inhibitors was subjected to quantitative structure–activity relationship (QSAR) analysis. For building the regression models two different variable selection approaches namely, genetic function approximation (GFA) and sequential multiple linear regression (SQ-MLR) were used and compared to predict the HIV-1IN inhibition activity. Based on prediction, the best validation model for 3′ processing inhibition activity with squared correlation coefficient (r^2) = 0.8965, cross validated correlation coefficient (Q^2) = 0.8307 and external prediction ability pred_r^2 = 0.5400 showed that Henry's law Constant (HLC), Partition Coefficient (PC) and Dipole moment-Z component (D3) were the positive contributors, whereas for integration inhibition activity, parameters r^2 = 0.8904, Q^2 = 0.8174 and pred_r^2 = 0.7159 showed HLC, Logarithm of Partition Coefficient (Log P) and Dipole moment-Y component (D2) contributed positively to the activity. The binding mode pattern of the compounds to the binding site of integrase enzyme was confirmed by docking studies. The results of the present study may be useful for designing more potent HIV-1IN inhibitors.

1. Introduction

Acquired immunodeficiency syndrome caused by HIV leads to life-threatening opportunistic infections, malignancies, functional impairment of the immune system and subsequently destroys the body's ability to fight against infections (Kanazawa and Matija Peterlin, 2001). Since first reported in 1981 (Gottlieb et al., 1981), it has spread rapidly through the human population and became one of the most devastating diseases faced by mankind. According to Joint United Nations Programme on HIV/AIDS, an estimated 34 million people were living with HIV worldwide at the end of 2010 (UNAIDS, 2011).

* Corresponding author. Tel.: +91 731 2454095.
E-mail addresses: vksrivastav@sgsits.ac.in (V.K. Srivastav), mtiwari@sgsits.ac.in, meenatiwari2004@yahoo.co.in (M. Tiwari).
[1] Tel.: +91 9827278138.

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1878-5352 © 2013 Production and hosting by Elsevier B.V. on behalf of King Saud University.
http://dx.doi.org/10.1016/j.arabjc.2013.01.015

Currently, the majority of available drugs for HIV treatment are Reverse Transcriptase (RT) and Protease (PT) Inhibitors. Rapid development of resistance to these inhibitors and toxicity are the major problems associated with the current therapy (Erickson and Burt, 1996). Therefore, the development of new anti-HIV agents with varied structure and mechanism of action is being focused. HIV-1IN is a very attractive and unexplored target to develop new anti-HIV drugs as it plays a vital role in replication cycle, has no cellular counterpart (Goldgur et al., 1999) and only one FDA (Food and Drug Administration) approved drug raltegravir is being used in clinical practice (Summa et al., 2008). Several reports have appeared on HIV-1IN inhibitory potency of variety of compounds. Examples of these compounds include lignanolides (Eich et al., 1996), curcumins (Mazumder et al., 1995), aurintricarboxylic acids (Cushman and Sherman, 1992), dicaffeoyl quinic acids and analogues (Robinson et al., 1996a,b), diaryl sulfones (Mazumder et al., 1996) and G-rich oligonucleotides (Hansch, 1969). These inhibitors may be described in general as consisting of two aryl units, at least one of which contains the 1,2-dihydroxy (catechol) pattern, separated by an appropriate linker. The utility of the catechol-containing inhibitors is significantly diminished by cytotoxicity due to *in situ* oxidation of the catechol moiety to reactive quinone species (Kostova et al., 2006). In an attempt to develop new inhibitors Zhao et al., synthesized a series of coumarin derivatives which do not contain catechol functionality but possess good HIV-1IN inhibition activity.

The QSAR approach helps to correlate the biological activity of a series of compounds with the calculated molecular properties in terms of descriptors (Hansch et al., 2001). It saves resources as well as speeds up the process of development of new molecules. Several QSAR Studies have been performed by other authors, which provide valuable insights in design and development of HIV-1IN inhibitors (Yuan and Parrill, 2002; Cheng et al., 2010; Kaushik et al., 2011). As a part of ongoing effort the present work is aimed to derive some statistically significant QSAR models for coumarin derivatives to correlate anti-HIV-1IN activity to its physicochemical properties. Docking study is also performed to a newly identified pocket right behind catalytic core domain (CCD) helix 4 (Rhodes et al., 2011) in the IN enzyme to understand the binding mode pattern of the compounds. The results obtained may contribute to further design and development of novel antiretroviral agents.

2. Materials and methods

2.1. Dataset

A dataset of coumarin derivatives containing 26 compounds with well defined activity (Zhao et al., 1997), was selected for QSAR study. The compounds which do not have well defined activity were excluded from dataset. The biological activity data in the form of IC_{50} (molar concentration of the drug leading to 50% inhibition of enzyme Integrase) value in μm (micromoles) were converted into negative logarithmic dose in moles (pIC_{50}) for QSAR Analysis (Table 1).

2.2. Molecular modeling and generation of molecular descriptors

The molecular modeling study was performed using CS ChemOffice version 8.0 software (Mendelsohn, 2004). Structure of all the compounds was drawn using ChemDraw Ultra module of the program and transferred to Chem3D Ultra module to create the three-dimensional (3D) structure. These structures were then subjected to energy minimization using molecular mechanics (MM2) server until the root-mean square (RMS) gradient value became smaller than 0.1 kcal/mol Å. Energy minimized molecules were subjected to reoptimization via Austin model-1 (AM1) method using the restricted closed-shell wave function of the molecular orbital package (MOPAC) module until the RMS gradient attained a value of 0.0001 kcal/mol Å. The geometry optimization of the lowest energy structure was carried out using EF (Eigenvector Following) routine. Most stable structure for each compound was generated and used for calculating various physicochemical parameters like thermodynamic, steric and electronic descriptors (Table S1 in Supplementary material).

2.3. Variable selection and model generation

Although many molecular descriptors are available, only a subset of them is statistically significant in terms of correlation with biological activity. Thus, it is very important to address the variable selection method for deriving the optimal QSAR model. GFA (Rogers and Hopfinger, 1994) and SQ-MLR approaches were adopted to select the best possible variables as well as for the generation of QSAR models.

2.3.1. GFA method

GFA algorithm is a search method to find exact or approximate solutions to optimization and search problems. GFA is conceived from (1) genetic algorithm and (2) Friedman's multivariate adaptive regression splines (MARS) algorithm. The following steps were performed: (1) initial population of equations were generated by random number of descriptors, (2) pairs from the population of equations were chosen at random, crossovers were performed and progeny equations were generated, (3) the fitness of each progeny equation was assessed by lack of fit (LOF) score that automatically penalizes models with too many features. A distinctive feature of GFA is that it generates a population of equations rather than a single equation as do most other statistical methods. The range of variations in this population gives added information on the quality of fit and importance of the descriptors. The fitness function, i.e., lack-of-fit is calculated by

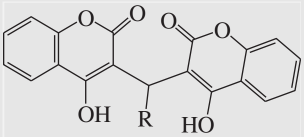
$$\mathrm{LOF} = \frac{\mathrm{LSE}}{\left(1 - \frac{C+dp}{M}\right)^2}$$

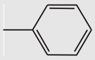
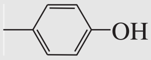
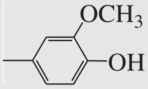
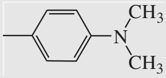
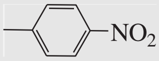
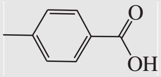
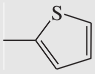
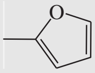
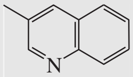
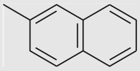
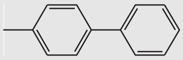
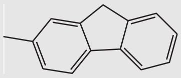
where c is the number of basis functions, d is the smoothing parameter, M is the number of samples in the training set, LSE is the least square error and p is the total number of features contained in all basis functions. Selected descriptors are given in supplementary material (Table S2).

2.3.2. SQ-MLR

In order to establish a correlation between physicochemical parameters as independent variables and pIC_{50} as dependent variable employing sequential multiple linear regression analysis method, dataset was transferred to the statistical program VALSTAT (Gupta et al., 2004). In sequential multiple regression, the program searches all the permutations and combinations sequentially for the data set. The ± data within the parentheses are the standard deviation, associated with the

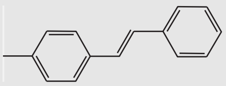
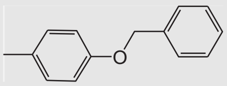
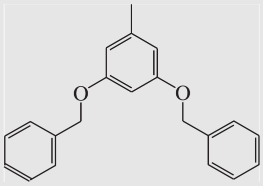
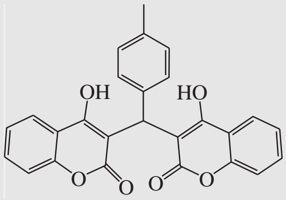
Table 1 Structure and biological activity of coumarin derivatives.

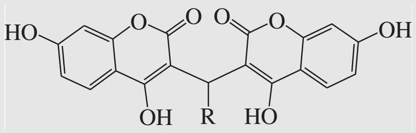


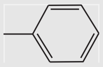
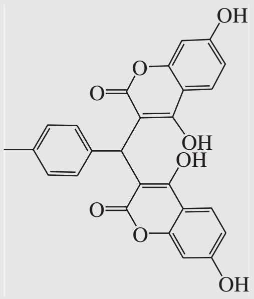
S. no.	*R*	IC_{50} (μM)		pIC_{50} (M)	
		3′ Processing	Integration	3′ Processing	Integration
1		43.4	38.8	4.3665	4.4111
2		128	74	3.8927	4.1307
3		177	76	3.7520	4.1191
4		88	50	4.0550	4.3010
5		50	12	4.3010	4.9208
6		48	49	4.3187	4.3098
7		100	148	4.0000	3.8297
8		34	112	4.4685	3.9507
9		29	14	4.5376	4.8538
10		19	9.0	4.7212	5.0457
11		14	13	4.8538	4.886
12		10	7.8	5.0000	5.1079

Please cite this article in press as: Srivastav, V.K., Tiwari, M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. Arabian Journal of Chemistry (2013), http://dx.doi.org/10.1016/j.arabjc.2013.01.015

Table 1 Continued.

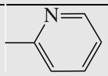
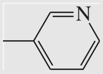
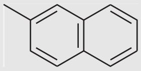
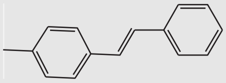
S. no.	*R*	IC_{50} (μM)		pIC_{50} (M)	
		3′ Processing	Integration	3′ Processing	Integration
13		5.5	3.7	5.2596	5.4317
14		8.5	14	5.0705	4.8538
15		8.0	3.7	5.0969	5.4317
16		1.5	0.8	5.6989	6.0969

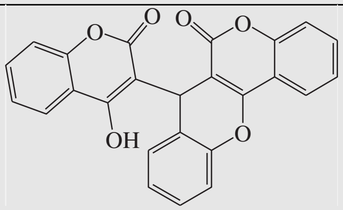
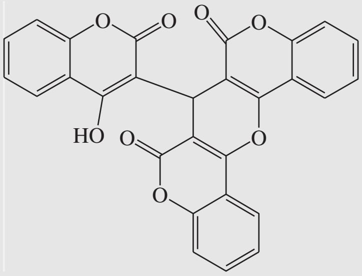
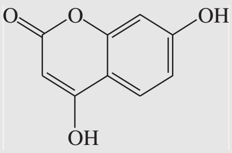


S. no.	R	IC_{50} (μM)		pIC_{50} (M)	
		3′ Processing	Integration	3′ Processing	Integration
17	H	46.3	44.9	4.3344	4.3477
18		17.2	22.2	4.7644	4.6536
19		0.37	0.33	6.4317	6.4814

coefficient of descriptors in regression equations. The best model was selected from the various statistically significant equations on the basis of the r^2, the standard error of estimate (SEE), the sequential Fischer test (*F*), the bootstrapping

Table 1 Continued.

S. no.	*R*	IC_{50} (μM)		pIC_{50} (M)	
		3′ Processing	Integration	3′ Processing	Integration
20		84	49	4.0757	4.3098
21		62	25	4.2076	4.6020
22		4.2	3.55	5.3767	5.4497
23		7.0	1.8	5.1549	5.7447

S. no.	Compound	IC_{50} (μM)		pIC_{50} (M)	
		3′ Processing	Integration	3′-Processing	Integration
24		121	122	3.9172	3.9136
25		35.7	22.5	4.4473	4.6478
26		300	325	3.5228	3.4881

squared correlation coefficient (r^2_{bs}), the bootstrapping standard deviation (S_{bs}), q^2 using leave-one-out method, significance level (P) on the basis of chance statistics (evaluated as the ratio of the equivalent regression equations to the total number of randomized sets; a chance value of 0.001 corresponds to 0.1% chance of fortuitous correlation) and outliers (on the basis of *Z*-score value).

Selection of training and test sets is one of the most important steps in QSAR analysis. There must be sufficient structural diversity which could cover the complete range of

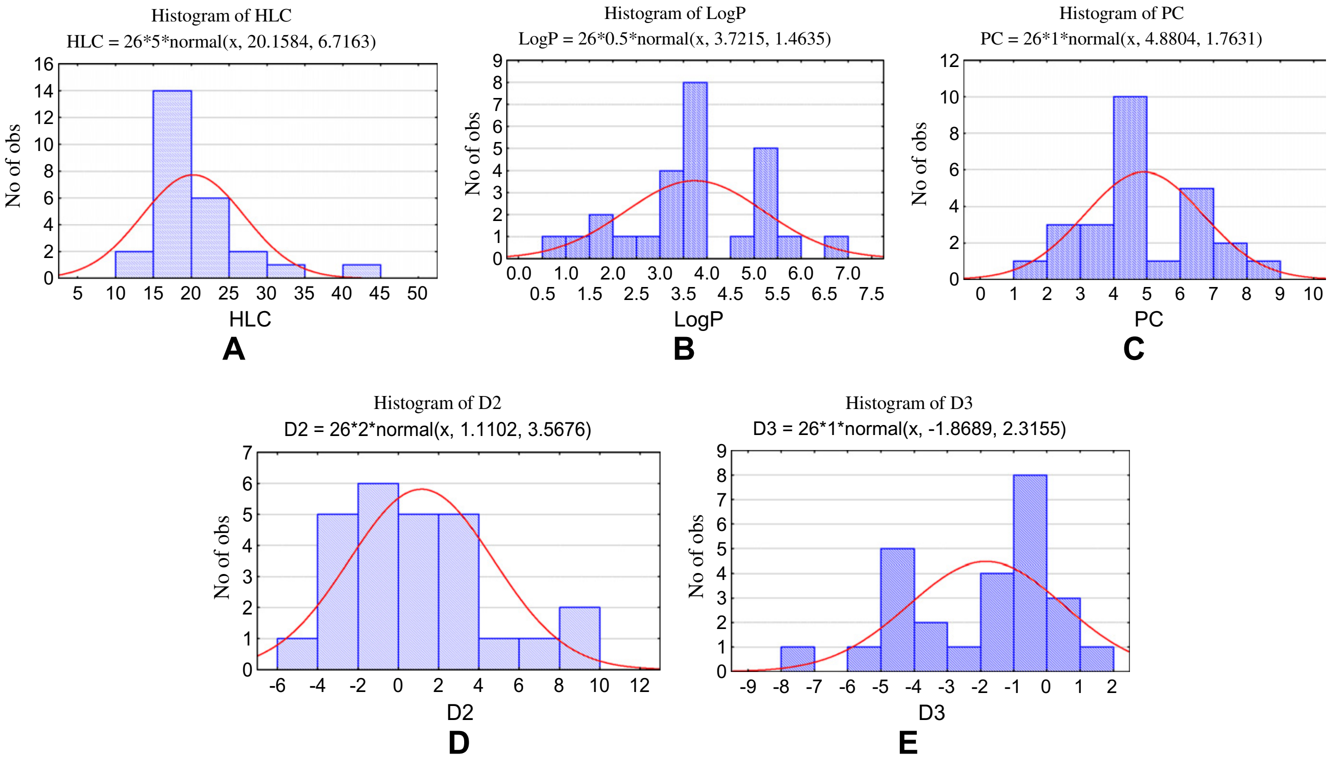


Figure 1 Normal distribution curves of individual descriptors used in the best validation models.

variation in biological activity. The dataset was divided into a training set of 18 compounds (**18, 19, 4, 17, 20, 6, 12, 22, 1, 13, 25, 16, 7, 11, 21, 26, 9,** and **5**) and test set of 8 compounds (**10, 8, 2, 3, 23, 24, 14,** and **15**) by straightforward random selection method (70:30 ratio) through activity sampling automatically by VALSTAT. After selection of training set and test set compounds, the dataset was further subjected to MLR analysis to predict the HIV-1IN inhibition activity of test set compounds using STATSTICA software (Statsoft, 2001). The descriptors selected for model generation are given in supplementary material (Table S2). The reliability of selected variables was checked by their normal distribution behavior. The distribution curves of individual descriptors used in the best validation models are given in Fig. 1.

2.4. Validation of QSAR models

The QSAR models were developed by GFA and SQ-MLR methods and evaluated using the following statistical parameters: n (the number of compounds in regression), correlation coefficient (r), r^2, F, Kubinyi function (FIT), variance, SEE, P, r^2_{bs}, and S_{bs} for statistical significance. 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. Kubinyi function (FIT) is a closely related statistical parameter to F-test. The greater the FIT value the better the linear equation. The validation parameters calculated were Q^2, standard deviation of sum of square of difference between predicted and observed values (S_{PRESS}) and standard deviation of error of prediction (S_{DEP}). For reliability of the model, probable error of correlation (PE) was also calculated. If the value of correlation coefficient (r) is more than six times of PE then the expression is good and reliable (Kumar et al., 2011). Finally, the derived QSAR models were used for the prediction of the biological activity of the compounds in the test set and pred_r^2 was calculated for evaluating the prediction ability of the models. A QSAR model is predictive if the following conditions are satisfied (Golbraikh and Tropsha, 2002a).

$$r^2 > 0.6, \quad q^2 > 0.6 \quad \text{and} \quad \text{pred\_r}^2 > 0.5.$$

Internal validation was carried out using ‘leave-one-out’ (q^2, LOO) method (Cramer et al., 1988). The cross-validated coefficient (q^2) was calculated using the following equation:

$$q^2 = 1 - \frac{\sum(y_{\text{obs}} - y_{\text{pred}})^2}{(\sum y_{\text{obs}} - y_{\text{mean}})^2}$$

where y_{obs} and y_{pred} are the observed and predicted activity of the training set, respectively, and y_{mean} is the average activity of all molecules in the training set. However, a high q^2 value does not necessarily give a suitable representation of the real predictive power of the model, so an external validation was also carried out. The external predictive ability of the selected models was calculated using the following equation:

$$\text{pred\_}r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2}$$

where y_i and $\hat{y}_i$ are the observed and predicted activity of the ith molecule in the test set, respectively, and y_{mean} is the average activity of all molecules in the training set.

The robustness of QSAR models was also checked by Y-randomization test. In this technique, the dependent variable (biological activity) is randomly shuffled and a new QSAR

Please cite this article in press as: Srivastav, V.K., Tiwari, M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. Arabian Journal of Chemistry (2013), http://dx.doi.org/10.1016/j.arabjc.2013.01.015

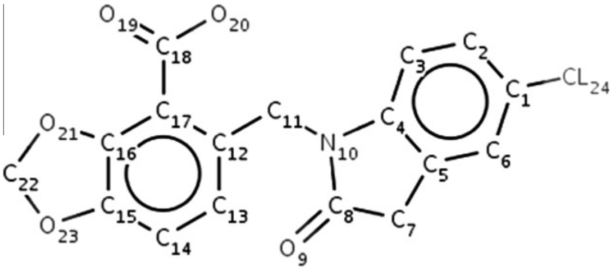


Figure 2 Structure of reference ligand.

model is developed using the original independent variable. The new QSAR models (after several trials) are expected to have low r^2 and Q^2 values, if the reverse happens then an acceptable QSAR model cannot be obtained for the specific modeling method and data. This technique ensures the robustness of a QSAR model (Tropsha et al., 2003; Wold et al., 2008).

The QSAR models developed using different statistical tools can be applied to define its applicability domain for different datasets. For this purpose, an external data set of 50 compounds (Sharma et al., 2011) with well defined HIV-1IN inhibition activity was selected which was different from the current series, and used as external test set for the prediction of their HIV-1IN inhibition activity (Table S3 in Supplementary material).

2.5. Docking study

To date, no full strength structure of HIV-1IN is available to elucidate the spatial arrangement of its three domains: N-terminal (NTD), catalytic core (CCD) and C-terminal (CTD). The relatively rapid emergence of resistance against raltegravir necessitates sustained research not only on novel INSTIs (IN strand transfer inhibitors) but also on entirely novel inhibitor classes. In recent years significant progress has been witnessed in the area of the development of allosterically targeted IN inhibitors. It presents an extremely advantageous approach for the discovery of compounds effective against IN strand-transfer drug-resistant viral strains (Al-Mawsawi and Neamati, 2011).

Docking was performed using GOLD software package version 4.0 (Genetic Optimization for Ligand Docking) (Jones et al., 1997). 3D co-crystallized structure of HIV-1IN was taken from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB ID: 3NF7). 3NF7 is a new site in integrase as a valid region for the structure-based design of allosteric integrase inhibitors which has been identified using a structure-based design process (Rhodes et al., 2011). Structure of all the compounds was built and energy minimized using Maestro module of Schrodinger Suit (Maestro, 2009). Feasible and unique conformations for each compound were generated over an energy range of 20 kcal/mol. Processing of the protein was done by adding hydrogen atoms to assign appropriate ionization states to both acidic and basic amino acid residues and removing remaining water molecules from protein. Binding site was defined by studying the interaction of reference ligand to the active residues in the cavity. The key characteristic of a good docking program is its ability to reproduce the experimental binding modes of ligands. To test this, ligand [*5-{(5-chloro-2-oxo-2, 3-dihydro-1H-indol-1- yl) methyl}-1, 3-benzodioxole-4-carboxylic acid*] was taken (as reference ligand) out of the X-ray structure of its protein–ligand complex (3NF7) and docked back into its binding site (Fig. 2).

The docked binding mode was then compared with the experimental binding mode, and a root-mean-square deviation (RMSD) between the two was calculated. The prediction of a binding mode is considered successful if the RMSD is below a certain value usually 2.0 Å (Verdonk et al., 2003). The exact superimposition was obtained with a RMSD value 0.4505 which confirmed that prediction of the binding mode was successful. Compounds from the dataset were docked in the previously defined cavity and interaction points as well as Goldscore fitness was compared with the reference ligand. Raltegravir was also included in the docking study to compare its binding orientation in active site with the reference ligand and compounds of the series. The redocked conformation of the

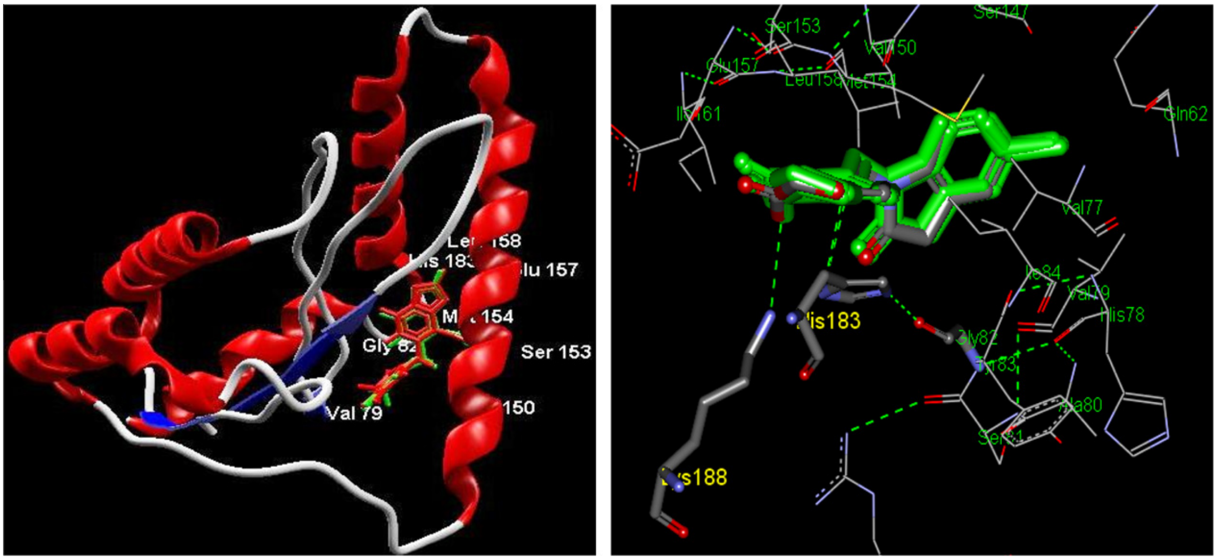


Figure 3 (a) Superimposition of the reference ligand. (b) Superimposition and binding of reference ligand to active residues in the cavity of HIV-1IN.

Table 2 Validation parameters of generated QSAR models.

Validation parameters	r^2_{bs}	S_{bs}	PE	P	Q^2	S_{press}	S_{DEP}	Pred_r^2	LOF
Model 2	0.8823	0.0965	0.0159	< 0.001	0.5353	0.5297	0.4671	0.4981	0.2675
Model 4	0.9045	0.0793	0.0162	< 0.001	0.8307	0.3184	0.2808	0.5400	0.2701
Model 6	0.9038	0.1024	0.0117	< 0.001	0.8705	0.2784	0.2456	0.5262	0.3376
Model 8	0.9050	0.0742	0.0172	< 0.001	0.8174	0.3576	0.3154	*0.7159*	0.3258

reference ligand and binding to the active residues in the binding site are shown in Fig. 3.

3. Results and discussion

3.1. QSAR study

The model generated for 3′ processing inhibition activity by GFA algorithm was Model 1.

3.1.1. Model 1

$\text{Log}(1/\text{IC}_{50}) = [3.2733(\pm 0.3430)] + \text{PMIZ}\ [8.2353(\pm 2.5328)] + \text{NVDW}\ [0.0161(\pm 0.0327)] + \text{TE}\ [-0.0248\ (\pm 0.0186)]$

$n = 26$, $r = 0.9046$, $r^2 = 0.8066$, Variance $= 0.0804$, SEE $= 0.2936$, PE $= 0.0213$, $F = 34.5544$, FIT $= 3.0089$, $r^2_{bs} = 0.8089$, $S_{bs} = 0.1257$, $P < 0.001$, $Q^2 = 0.7076$, $S_{press} = 0.3508$, $S_{DEP} = 0.3243$ LOF $= 0.3392$.

Validation was performed by dividing dataset into training set and test set. The best model generated for 3′ processing inhibition activity using GFA method was Model 2.

3.1.2. Model 2

$\text{Log}(1/\text{IC}_{50}) = [3.3853(\pm 0.3483)] + \text{PMIZ}\ [7.7998\ (\pm 2.4799)] + \text{SE}\ [-0.0092(\pm 0.0254)] + \text{TE}\ [-0.0259(\pm 0.0170)]$

$n = 18$, $r = 0.9478$, $r^2 = 0.8983$, Variance $= 0.0613$, SEE $= 0.2477$, PE $= 0.0159$, $F = 39.2407$, FIT $= 4.0823$, $r^2_{bs} = 0.8823$, $S_{bs} = 0.0965$, $P = 0.001$, $Q^2 = 0.5353$, $S_{press} = 0.5297$, $S_{DEP} = 0.4671$

SQ-MLR analysis resulted in several significant models with respect to inhibition of 3′ processing and integration activity, respectively. Model 3 was selected for 3′ processing inhibition activity.

3.1.3. Model 3

$\text{Log}(1/\text{IC}_{50}) = [2.4610(\pm 0.5291)] + \text{HLC}\ [0.0524(\pm 0.0194)] + \text{PC}\ [0.2376(\pm 0.0698)] + \text{D3}\ [0.0410(\pm 0.0561)]$

$n = 26$, $r = 0.9095$, $r^2 = 0.8273$, Variance $= 0.0856$, SEE $= 0.2926$, PE $= 0.0225$, $F = 35.1433$, FIT $= 3.0122$, $r^2_{bs} = 0.8048$, $S_{bs} = 0.1158$, $P < 0.001$, $Q^2 = 0.7569$, $S_{press} = 0.3472$, $S_{DEP} = 0.3193$, LOF $= 0.4072$.

The best Validation model for 3′ processing inhibition activity was Model 4.

3.1.4. Model 4

$\text{Log}(1/\text{IC}_{50}) = [2.4937(\pm 0.5187)] + \text{HLC}\ [0.0516(\pm 0.0190)] + \text{PC}\ [0.2456(\pm 0.0733)] + \text{D3}\ [0.0392(\pm 0.0268)]$

$n = 18$, $r = 0.9468$, $r^2 = 0.8965$, Variance $= 0.0620$, SEE $= 0.2489$, $F = 40.4276$, FIT $= 4.4919$, $r^2_{bs} = 0.9045$, $S_{bs} = 0.0793$, $P = 0.001$, $Q^2 = 0.8307$, $S_{press} = 0.3184$, $S_{DEP} = 0.2808$.

Based on the statistical significance and validation parameters a comparison was done between the validation models (Model 2 and Model 4 for 3′ processing activity) generated by GFA and MLR methods (Table 2).

Model 2 showed lower Q^2 and pred_r^2 values than Model 4 which meaned that prediction ability of Model 4 was much better. Bootstrapping analysis was performed to access the robustness and statistical confidence. Higher value of r^2_{bs} and lower value of S_{bs}, S_{press} and S_{DEP} of Model 4 in comparison to Model 2 revealed that Model 4 was robust and promising. In the developed Model the value of coefficient of correlation was significantly higher than the value of PE (0.097) supporting reliability and goodness. Based on the above results Model 4 was considered as the best validation model for 3′ processing inhibition activity. The accuracy of the Model 4 was ascertained by correlation coefficient ($r = 0.9468$), statistical significance more than 99% (against tabulated value $F = 40.4276$) and low standard deviation (0.2489). The model shows that thermodynamic parameter (HLC and PC) and electronic parameter (D3) showed positive contribution. The correlation matrix between the physicochemical parameters and the biological activity is presented in Table S4 (supplementary material). HLC is a quantitative expression of the compound's partitioning nature between two phases. It is evident from Table S2 that most potent compound of the series has the highest value for HLC = 44.4244 and least active compound has the lowest value for HLC = 11.0686. PC can be calculated using an atom based approach and represents the hydrophobicity of the compounds. The positive contribution of PC confirmed the hydrophobic binding site of integrase protein and suggested that substitution with lipophilic groups favors anti-HIV-1IN activity. The measurement of dipole moment helps in distinguishing between polar and non-polar molecules, as it encodes information about the charge distribution in molecules. It is particularly important in modeling solvation properties of compounds which depend on solute/solvent interactions and in fact are frequently used to represent the dipolarity/polarizability term in linear solvation energy relationships. Moreover, it can be used to model the polar interactions which contribute to the determination of the lipophilicity

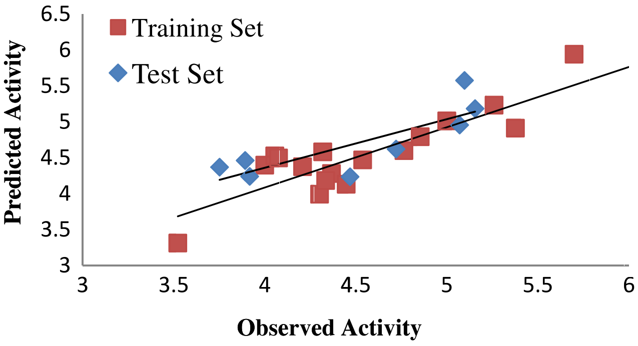


Figure 4 Scatter Plot between the observed and predicted activity of training and test set (3′ processing inhibition activity).

Please cite this article in press as: Srivastav, V.K., Tiwari, M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. Arabian Journal of Chemistry (2013), http://dx.doi.org/10.1016/j.arabjc.2013.01.015

of compounds (Todeschini and Consonni, 2008). Non-polar molecules have zero dipole moment while polar molecules have some value of dipole moment. The dipole moment of compound nos. **16** and **19** was zero because the groups on either side of linker were same. The robustness of the model was justified by the magnitude of bootstrapping r^2 (0.8982), which was near to the conventional r^2 (0.8965). The internal validation parameter of the model ($q^2 = 0.8307$) was also good. The scatter plot of observed activity versus predicted activity is shown in Fig. 4.

The model selected for integration inhibition activity by GFA was Model 5.

3.1.5. Model 5

$\text{Log}(1/\text{IC}_{50}) = [3.1965(\pm 0.4617)]$ + PMIZ $[8.9258\ (\pm 2.9132)]$ + ce:hsp sp="0.25"/>NVDW $[-0.0195(\pm 0.0170)]$ + TE $[-3.0810(\pm 7.4373)]$.

$n = 26$, $r = 0.9217$, $r^2 = 0.8496$, Variance $= 0.0745$, SEE $= 0.2731$, PE $= 0.0196$, $F = 41.4302$, FIT $= 3.5511$, $r^2_{\text{bs}} = 0.8402$, $S_{\text{bs}} = 0.1084$, $P < 0.001$, $Q^2 = 0.8169$, $S_{\text{press}} = 0.3013$, $S_{\text{DEP}} = 0.2771$, LOF $= 0.3063$.

The best validation model resulted by dividing test set and training set using GFA method was Model 6.

3.1.6. Model 6

$\text{Log}(1/\text{IC}_{50}) = [3.2314(\pm 0.3253)]$ + PMIZ $[8.1759(\pm 2.2220)]$ + VDWE $[-0.0231(\pm 0.0156)]$ + TC $[0.0362(\pm 0.0558)]$.

$n = 18$, $r = 0.9618$, $r^2 = 0.9251$, Variance $= 0.0448$, SEE $= 0.2117$, PE $= 0.0117$, $F = 35.6723$, FIT $= 6.408$, $r^2_{\text{bs}} = 0.9058$, $S_{\text{bs}} = 0.1024$, $P < 0.001$, $Q^2 = 0.8705$, $S_{\text{press}} = 0.2784$, $S_{\text{DEP}} = 0.2456$.

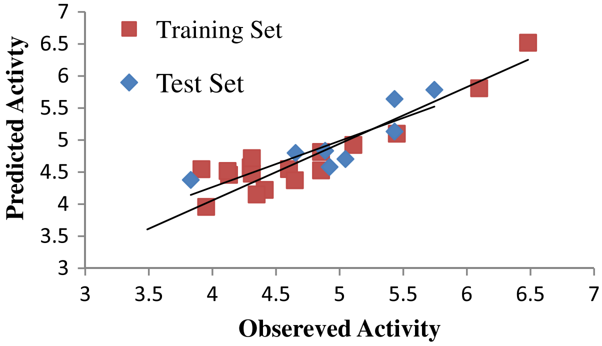


Figure 5 Scatter plot between the observed and predicted activity of training and test set (Integration inhibition activity).

The model selected for integration inhibition activity by MLR was Model 4.

3.1.7. Model 7

$\text{Log}(1/\text{IC}_{50}) = [2.1070(\pm 0.4586)]$ + HLC $[0.0637(\pm 0.0169)]$ + PC $[0.2675(\pm 0.0655)]$ + D2 $[0.0409(\pm 0.0321)]$

$n = 26$, $r = 0.9349$, $r^2 = 0.8740$, Variance $= 0.0742$, SEE $= 0.2724$, $F = 50.9052$, FIT $= 4.3633$, PE $= 0.0164$, $P < 0.001$, $Q^2 = 0.8346$, $S_{\text{press}} = 0.3122$, $S_{\text{DEP}} = 0.2871$. LOF $= 0.3767$.

Table 3 Observed, calculated and predicted activity (3′ processing and integration inhibition) of Coumarin derivatives.

S. no.	Obs.[a]	Obs.[b]	Model 3		Mode7		Model 4		Model 8	
			Cal.	Pred.	Cal.	Pred.	Cal.	Pred.	Cal.	Pred.
1	4.3665	4.4111	4.2241	4.2096	4.2761	4.2620	4.2919	4.2809	4.2471	4.2224
2	3.8927	4.1307	4.4084	4.4358	4.3617	4.3823	*	4.4620	4.4261	4.4572
3	3.7520	4.1191	4.3117	4.3756	4.4045	4.4308	*	4.3687	4.4809	4.5176
4	4.0555	4.3010	4.3799	4.4235	4.5091	4.5252	4.4483	4.5253	4.5455	4.5738
5	4.3010	4.9208	4.0406	3.9319	4.5946	4.5304	4.1121	3.9925	*	4.5800
6	4.3187	4.3098	4.5089	4.5169	4.5152	4.5318	4.5668	4.5818	4.4603	4.4762
7	4.0000	3.8297	4.2954	4.3192	4.3618	4.3952	4.3540	4.3969	*	4.3781
8	4.4685	3.9507	4.1810	4.1528	4.2358	4.2569	*	4.2355	3.9566	3.9574
9	4.5376	4.8538	4.4140	4.4034	4.4987	4.4701	4.4775	4.4700	4.5593	4.5272
10	4.7212	5.0457	4.5436	4.5207	4.7478	4.7249	*	4.6207	*	4.7033
11	4.8538	4.8860	4.7228	4.7022	4.8451	4.8393	4.8053	4.7928	*	4.8327
12	5.0000	5.1079	4.9356	4.9250	4.9861	4.9729	5.0102	5.0133	4.9669	4.9237
13	5.2596	5.4317	5.1638	5.1340	5.2185	5.1829	5.2426	5.2320	*	5.1335
14	5.0705	4.8538	4.8816	4.8622	4.7813	4.7670	*	4.9553	4.8257	4.8142
15	5.0969	5.4317	5.4911	5.6154	5.6125	5.6562	*	5.5760	*	5.6414
16	5.6989	6.0969	5.8067	5.8337	6.0151	5.9956	5.8769	5.9412	5.8797	5.8079
17	4.3344	4.3477	4.1659	4.1426	4.2639	4.2523	4.2042	4.1826	4.1886	4.1513
18	4.7644	4.6536	4.5773	4.5540	4.7185	4.7248	4.6203	4.5996	*	4.7988
19	6.4317	6.4814	6.2277	5.9569	6.5570	6.6562	6.2760	6.0007	6.4959	6.5186
20	4.0757	4.3098	4.3718	4.4485	4.4319	4.4594	4.4007	4.4974	4.6496	4.7175
21	4.2076	4.6020	4.3077	4.3302	4.4521	4.4183	4.3394	4.3751	4.5585	4.5469
22	5.3767	5.4497	4.9213	4.8672	5.0553	5.0289	4.9725	4.9122	5.1408	5.0990
23	5.1549	5.7447	5.1118	5.1054	5.7785	5.7953	*	5.1837	*	5.7845
24	3.9172	3.9136	4.1741	4.2001	4.4200	4.5109	*	4.2401	4.3597	4.5454
25	4.4473	4.6478	4.1010	4.0751	4.5395	4.5003	4.1607	4.1305	4.5157	4.3725
26	3.5228	3.4881	3.3570	3.2834	3.1369	3.0015	3.3921	3.3128	3.1171	2.8922

[a] Obs.

[b] Obs are the observed activity against 3′ processing and integration, respectively.

* Represents the test set compounds.

Table 4 Predicted activity and Goldscore fitness of external data set.

S. no.	3′ Processing activity (pIC_{50})			Integration activity (pIC_{50})			Goldscore fitness
	Obs.	Pred.	Residual	Obs.	Pred.	Residual	
1	4.1249	4.5308	−0.4058	4.6020	4.4224	0.1796	56.8326
2	4.2839	4.6846	−0.4006	4.8538	4.4916	0.3622	56.3636
3	4.1307	4.5332	−0.4024	4.4948	4.3735	0.1213	54.8883
4	4.3467	4.6592	−0.3124	5.0457	4.4632	0.5824	54.0343
5	4.0915	4.6671	−0.5756	4.9208	4.4269	0.4938	56.3896
6	4.9586	4.7129	0.2456	5.3010	4.5242	0.7767	54.7906
7	4.1611	4.7895	−0.6283	4.4559	4.6995	−0.2436	60.4016
8	4.5528	4.8112	−0.2583	5.1549	4.6482	0.5066	58.3345
9	4.2365	4.7957	−0.5591	4.7212	4.6787	0.0425	54.1915
10	4.2076	4.5245	−0.3169	4.4685	4.4130	0.0554	59.1822
11	4.0705	4.0705	4.3814	−0.3108	4.2924	4.2728	0.0196
12	4.3279	4.9320	−0.6041	4.7447	4.7248	0.0199	55.0176
13	4.2839	4.6509	−0.3669	4.7958	4.7449	0.0509	53.9539
14	4.5686	4.7695	−0.2009	4.8538	4.6177	0.2361	56.3071
15	4.6989	4.9911	−0.2921	4.8860	4.7968	0.0892	60.0161
16	4.6382	5.0155	−0.3773	5.4317	4.8281	0.6036	61.5000
17	4.0705	4.6188	−0.5482	4.2757	4.3225	−0.0468	57.7221
19	4.0000	4.7439	−0.7439	4.0362	4.5724	−0.5362	53.5698
20	4.1249	4.8786	−0.7536	4.6020	4.6166	−0.0145	59.8230
21	4.2596	4.9494	−0.6898	4.7695	4.6887	0.0808	54.2735
22	4.1135	4.9639	−0.8504	4.5228	4.7208	−0.1979	60.1604
23	4.0604	4.7615	−0.7011	4.3872	4.5059	−0.1187	56.5870
24	4.4436	4.4703	−0.0266	4.7695	4.2005	0.5690	55.4558
25	4.2839	4.6406	−0.3566	4.6197	4.4169	0.2028	53.0649
26	4.2596	4.5983	−0.3386	4.5228	4.4506	0.0722	52.9971
27	4.3010	4.6502	−0.3492	4.5376	4.4848	0.0527	55.4299
28	4.4814	4.7957	−0.3142	4.8239	4.5665	0.2573	56.6054
29	4.0861	4.8312	−0.7450	4.4089	4.5937	−0.1848	52.0521
33	4.6777	5.0294	−0.3516	5.0457	5.1204	−0.0747	67.2311
34	4.6020	5.2405	−0.6384	4.7447	5.1334	−0.3887	67.2029
35	4.6575	5.2437	−0.5861	4.7958	5.1445	−0.3486	76.7754
36	4.5228	4.4318	0.0910	4.7447	4.2283	0.5164	56.6828
37	4.0000	4.9108	−0.9108	4.2596	4.5060	−0.2463	63.1383
38	4.0362	4.8867	−0.8505	4.3372	4.8040	−0.4668	59.8100
39	4.2676	4.8141	−0.5465	4.2676	4.5032	−0.2356	61.4062
40	4.6382	4.6846	−0.0463	4.9586	4.4979	0.4606	60.8004
42	4.1023	4.7720	−0.6697	4.2006	4.4861	−0.2854	58.8186
44	4.3098	4.1725	0.1373	4.4948	3.8466	0.6482	56.9085
45	4.7212	4.7043	0.0168	4.7447	4.5646	0.1800	57.0653
46	4.1487	4.4963	−0.3476	4.2291	4.3251	−0.0959	52.7239
50	4.4436	4.5294	−0.0857	4.6777	4.1895	0.4882	53.0475

Model 8 was the validation model for integration inhibition activity.

3.1.8. Model 8

$\text{Log}(1/IC_{50}) = [2.1032(\pm 0.5361)] + \text{HLC}\ [0.0726(\pm 0.0190)] + \text{Log}\,P\ [0.3047(\pm 0.1135)] + \text{D2}\ [0.0346(\pm 0.0257)]$

$n = 18$, $r = 0.9436$, $r^2 = 0.8904$, Variance $= 0.0767$, SEE $= 0.2770$, $F = 37.9443$, FIT $= 4.2160$, $r^2_{bs} = 0.905$, $S_{bs} = 0.0742$, $P < 0.001$, $Q^2 = 0.8174$, $S_{press} = 0.3576$, $S_{DEP} = 0.3154$.

From Table 2 it is evident that Model 6 showed better value for Q^2 (0.8705) than Model 8 (0.8174) but a lower value for r^2_{bs}. A high value of Q^2 alone is an insufficient criterion for a QSAR model to be highly predictive (Golbraikh and Tropsha, 2002b). Based on prediction ability, Model 8 was selected as the best validation model for integration inhibition activity. Model 8 shows a good correlation between descriptors (HLC, Log P and D2) and integration inhibition activity. The correlation matrix between the physicochemical parameters and the biological activity is given in Table S5 (supplementary material). Correlation coefficient ($r = 0.9436$), squared correlation coefficient ($r^2 = 0.8904$), Low standard deviation value (0.2770) of the model and a statistical significance more than 99% (F value = 37.9443) demonstrate the accuracy of the model. Positive contribution of HLC, Log P and D2 indicated favorable hydrophobic interactions that were responsible for the enhancement of HIV-1IN inhibition activity. Log P is one criterion used in medicinal chemistry to assess the drug-likeness of a given molecule and used to calculate lipophilicity, a function of potency and Log P that evaluate the quality of designed compounds (Leeson and Springthorpe, 2007; Edwards and Price, 2010). The scatter plot between calculated and predicted activities of the training set and test set compounds is given in Fig. 5.

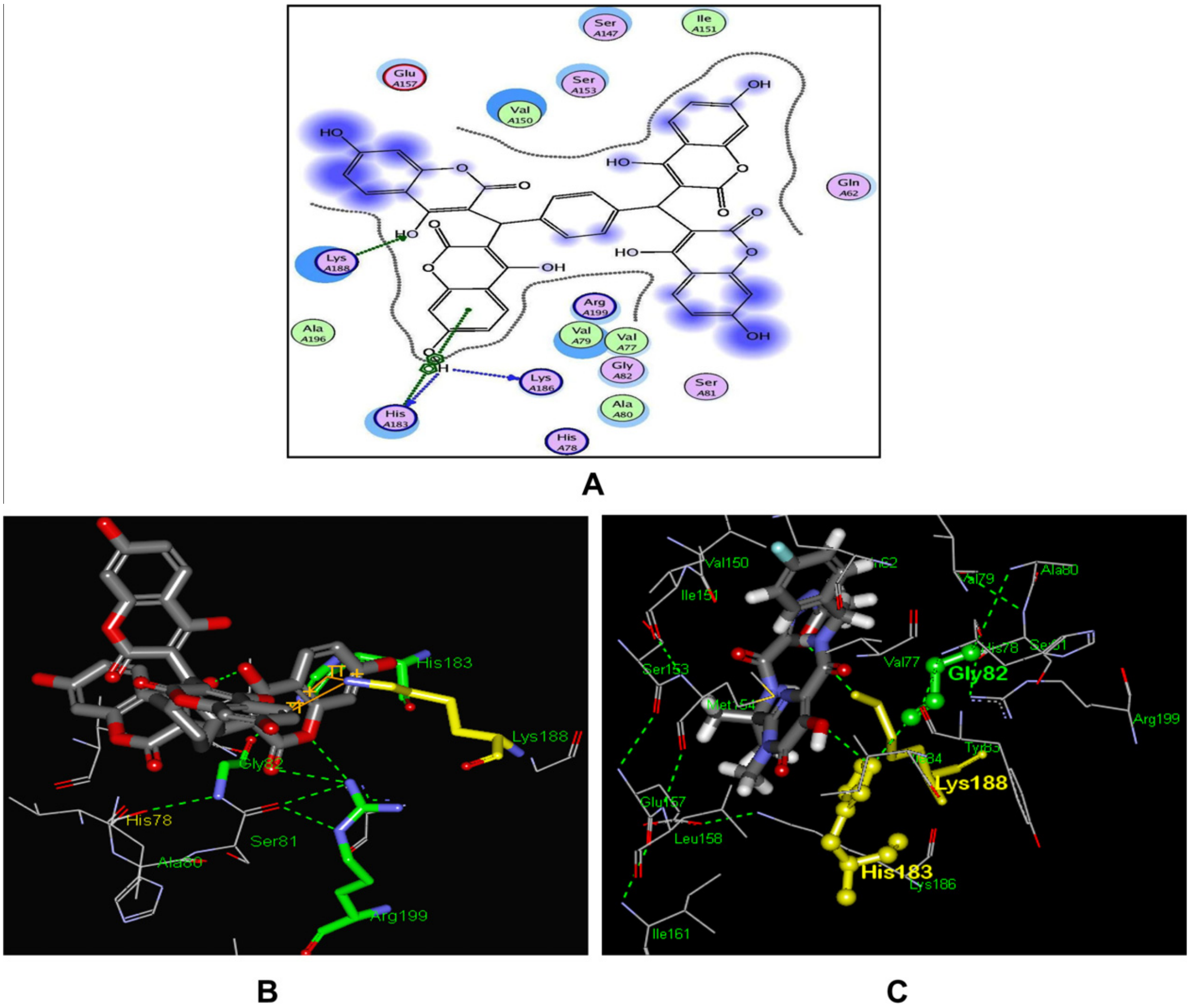


Figure 6 A. Docked pose of compound no.**19** in Binding site of 3NF7 [(green dots = Arene–Arene and H-Bond interaction), B. H-Bond Interaction of compound no.**19** (green dots = H-bond, red = oxygen, grey = carbon, blue = nitrogen) to His183 (green) and Lys188 (yellow) in the binding site. C. Interaction of the Raltegravir (green dots = H-bond, red = oxygen, grey = carbon, blue = nitrogen) with the active residues [His183 (yellow) and Lys188 (yellow)] of integrase binding site.

To confirm the robustness of the derived best validation models, a y-randomization test was performed by scrambling the experimental activity at 1000 random numbers of trial considering the same number and definition of descriptors. The results so obtained show that original model was not obtained due to a chance correlation. (Table S6 in Supplementary material). All the regression models were checked for the presence of outliers using *Z*-score method. There were no outliers present in any model. Low value for LOF for all the models suggested that selected models for both activities were robust. The predicted biological activities of training and test set molecules are given in Table 3.

From Table 3, it is evident that the predicted activities of all compounds in the training set and test set are in good agreement with their corresponding experimental activities. As the number of aromatic rings increases, the resonance energy per π electron decreases. As a result, larger polynuclear aromatic hydrocarbons have a tendency to undergo addition reaction to an internal ring to give more stable compounds. Compound nos. **16** and **19** in series have a large, highly complex but symmetrical structure consisting of a coumarin dimer containing the aryl substituent on the central linker methylene. As the two of the original four coumarin units were removed from compound 19, reduction in the potency resulted (compound no. **18**). Further if one more coumarin unit was removed from compound 18 along with the linker, significant reduction of potency resulted (compound no. **26**). Replacement of the central phenyl ring by more extended aromatic system having higher lipophilicity resulted in the enhancement of potency. Presence of nitrogen in compound nos. **9**, **20** and **21** reduces the aromaticity, has low value for Log *P* and PC and thus showed low biological activity. Dipole was also contributing positively to the activity to a small extent suggesting that the moiety, which increases the charge distribution over the molecules, is favorable for the activity. For better understanding the effect of dipole moment, DDE was also calculated which shows that potent compounds in the series (compound nos. 16 and 19) have higher DDE value (25.8412 and 23.4976, respectively) and least active compound (compound no. 26) has the least DDE value (1.3283). The results of the present

Table 5 Specific H-Bond interactions between coumarin derivatives and amino acids.

S. no.	Compound number	H-Bond interaction
1	**3**, **5**, **6**,	Arg199
2	**17**, **18**, **20**, **21** and **22**	Leu158
3	**24**	Met154
4	**22**	Lys188
5	**7** and **8**	Val150
6	**26**	Arg199, Lys186, Lys188 and Ala196

study confirm that suitable aromatic substitution with high lipophilicity is essential for HIV-1IN inhibition activity. QSAR study by other authors (Bansal et al., 2007; Kaushik et al., 2011) also resulted that lipophilicity is an important parameter for HIV-1IN inhibition activity. Further, the selected models were used to predict the HIV-1IN inhibition activity of external dataset. Pred_r^2 and residual values were calculated for both activities. The prediction attained from validation models was in agreement with experimental activity (residual values <0.7767) (Table 4).

3.2. Docking interpretation

The docked conformations showed that all ligands bind to the active residues in the predefined hydrophobic binding pocket. The location of the docked ligands agreed well with that of the docked reference ligand. The key residues in the binding pocket were His183, Met154, Gly82, Glu157, Leu158, Ile151, Val150, Val77 and Ile84. The reference ligand formed hydrogen bonding (H-bond) with His183 and Lys188 in the binding site. His183 further binds with Gly82 through H-bond. Most of compounds in the series showed H-bond interaction with His183 which confirmed that they interact with the integrase enzyme like the reference ligand. In case of docked poses of Raltegravir, H-bonds with His 183 and Lys188 were formed which further confirmed that these residues were important for the HIV-1IN inhibition activity. Compound nos. **16** and **19**, the most potent compound of the series showed a Goldscore fitness value of 72.1476 and 66.3396, respectively (Fig. 6).

Table 6 Goldscore fitness value of the coumarin derivatives.

S. no.	Goldscore fitness
1	60.5947
2	59.9817
3	58.7115
4	63.2896
5	60.4170
6	59.2247
7	58.8453
8	59.3408
9	73.3419
10	61.2063
11	74.8875
12	61.6435
13	76.5358
14	58.1679
15	62.6386
16	72.1476
17	60.8233
18	60.3536
19	66.3396
20	59.7796
21	62.7426
22	60.6169
23	75.4920
24	61.9995
25	59.7655
26	35.5097
Reference ligand	67.8933
Raltegravir	53.3109

Few compounds have Goldscore fitness more than potent compounds. This indicated that the compounds studied may differ in the exact relationship between structure and inhibition, through interactions with different subsets of amino acids in the binding pocket other than the active residues and through the presence of non-overlapping binding pockets. Interaction study is also important along with Goldscore fitness to validate the activity of compounds. The docked reference ligand was found to have H-Bond interaction between an oxygen atom (O_{23}) of the ligand and NH group of His183, O_{19} of ligand and NH_2 group of Lys188. Moreover, a π-cation and π–π interaction was also found in the docked conformation. His183 forms a π–π, and hydrophobic interaction with most of the compounds of the series. Compound no. **19** forms a π-cation interaction with Lys188 and a π–π interaction with His183. In order to explain the binding of these compounds, the H-bonding interactions with the other surrounding residues in the hydrophobic binding pocket were also investigated. Specific H-bond interactions between coumarin derivatives and amino acids are given in Table 5.

Strong H-bond interactions between the hydroxyl group of Compound 16 and an oxygen atom of Gly82 and Lys188 were formed. H-bond interactions between amino acids were also found in the binding pocket (His183:Gly82, His183:Ala179, Ile161:Glu157, Leu158:Met154, Lys188:Lys186, Arg199:-Ser195, Arg199:Ser81, Lys186:Glu157 and Met154:Ile151). The Goldscore fitness of the compounds is given in Table 6.

The external data set compounds also showed good Goldscore fitness (Table 4) and formed H-bond interaction with the active residues (His183 and Lys188) in the binding site. Studies on mutational effect of residues of HIV-1IN showed that mutations of His183, Lys188, and phenylalanine 185 had minor effects on the capacity of the mutated IN (Leavitt et al., 1996; Berthoux et al., 2007). From the docking studies it can be concluded that His183, Gly82, and Lys188, are the key residues of the integrase binding site. Raltegravir also interacts with these residues which further confirmed that binding to these residues is important for good anti HIV-1IN activity. The docking results agreed well with the observed biological activity data, which showed that the HIV-1IN inhibition activity of the series conforms to the docking results.

4. Conclusion

The present study showed that two coumarin units attached via an aryl linker were important for HIV-1IN inhibition activity. Removal of one coumarin unit resulted in lowering of activity (compound no. **26**). Through the iterative computational approach, it is possible to extract a simple and highly informative model, having a high degree of predictability for activity of coumarin derivatives against HIV-1IN. The generated QSAR models were very informative as they showed statistical significance more than 99% and good prediction ability. Further from the prediction ability of external dataset it is confirmed that the present QSAR study holds true for different sets of compounds. The results of this study suggested that substitution at the linker with substituted aromatic rings having good lipophilic characteristic was favorable for activity. Partition coefficient contributed positively for both activities, responsible for the hydrophobicity of the molecules. HLC, PC, Log *P*, and D (dipole moment) were found to be the important parameters for the HIV-1IN inhibition activity. The result of docking study revealed that compounds showed interaction with the active residues (His183, Lys188) in the binding site like the reference ligand. Goldscore fitness of the potent compounds (**16** and **19**) was comparable with the reference as well as with raltegravir. It can also be concluded that interaction study is a very important factor along with Goldscore fitness to understand the binding mode pattern of any compound in the binding site. Further from this study it is evident that 3NF7 can be used as a good model for rational drug design of HIV-1IN inhibitors.

Acknowledgements

Authors are thankful to the Director, SGSITS for providing state of art facilities to carry out research, University Grant Commission, New Delhi, for providing financial support and the referees of this paper for their thoughtful review and helpful suggestions for improving the clarity and quality of this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.arabjc.2013.01.015.

References

Al-Mawsawi, L.Q., Neamati, N., 2011. Chem. Med. Chem. 6, 228–241.

Bansal, R., Karthikeyan, C., Moorthy, H.N., Trivedi, P., 2007. ARKIVOC 15, 66–81.

Berthoux, L., Sebastian, S., Muesing, M.A., Luban, J., 2007. Virology 364, 227–236.

Cheng, Z., Zhang, Y., Fu, W., 2010. Eur. J. Med. Chem. 45, 3970–3980.

Cramer, R.D., Patterson, D.E., Bunce, J.D., 1988. J. Am. Chem. Soc. 110, 5959–5967.

Cushman, M., Sherman, P., 1992. Biochem. Biophys. Res. Commun. 185, 85–90.

Edwards, M.P., Price, D.A., 2010. Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks. In: John, E.M. (Ed.), Annu. Academic Press, Rep. Med, pp. 380–391.

Eich, E., Pertz, H., Kaloga, M., Schulz, J., Fesen, M.R., Mazumder, A., Pommier, Y., 1996. J. Med. Chem. 39, 86–95.

Erickson, J.W., Burt, S.K., 1996. Annu. Rev. Pharmacol. Toxicol. 36, 545–571.

Golbraikh, A., Tropsha, A., 2002a. J. Comput. Aided Mol. Des. 16, 357–369.

Golbraikh, A., Tropsha, A., 2002b. J. Mol. Graph. Model. 20, 269–276.

Goldgur, Y., Craigie, R., Cohen, G.H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., Davies, D.R., 1999. Proc. Natl. Acad. Sci. 96, 13040–13043.

Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A., Saxon, A., 1981. N. Engl. J. Med. 305, 1425–1431.

Gupta, A.K., Arockia Babu, M., Kaskhedikar, S.G., 2004. Indian J Pharm. Sci. 66, 396–402.

Hansch, C., 1969. Acc. Chem. Res. 2, 232–239.

Hansch, C., Kurup, A., Garg, R., Gao, H., 2001. Chem. Rev. 101, 619–672.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. J. Mol. Biol. 267, 727–748.

Kanazawa, S., Matija Peterlin, B., 2001. Microbes Infect. 3, 467–473.

Kaushik, S., Gupta, S.P., Sharma, P.K., Anwar, Z., 2011. Med. Chem. 7, 553–560.

Kostova, I., Raleva, S., Genova, P., Argirova, R., 2006. Bioinorg. Chem. Appl., 2006.

Kumar, S., Singh, V., Tiwari, M., 2011. Med. Chem. Res. 20, 1530–1541.

Leavitt, A.D., Robles, G., Alesandro, N., Varmus, H.E., 1996. J. Virol. 70, 721–728.

Leeson, P.D., Springthorpe, B., 2007. Nat. Rev. Drug. Discov. 6, 881–890.

Maestro, 2009. Version 9.0, Schrodinger, LLC, New York.

Mazumder, A., Raghavan, K., Weinstein, J., Kohn, K.W., Pommier, Y., 1995. Biochem. Pharmacol. 49, 1165–1170.

Mazumder, A., Neamati, N., Sommadossi, J.P., Gosselin, G., Schinazi, R.F., Imbach, J.L., Pommier, Y., 1996. Mol. Pharmacol. 49, 621–628.

Mendelsohn, L.D., 2004. ChemDraw 8 Ultra, Windows and Macintosh Versions. J. Chem. Inform. Comput. Sci. 44, 2225–2226.

Rhodes, D.I., Peat, T.S., Van de Graaf, N., Jeevarajah, D., Le, G., Jones, E.D., Smith, J., Coates, J.A.V., Winfield, L.J., Thienthong, N., Newman, J., Lucent, D., Ryan, J.H., Savage, G.P., Francis, C.L., Deadman, J.J., 2011. Antivir. Chem. Chemother. 21, 155–168.

Robinson, W.E., Cordeiro, M., Abdel-Malek, S., Jia, Q., Chow, S.A., Reinecke, M.G., Mitchell, W.M., 1996a. Mol. Pharmacol. 50, 846–855.

Robinson, W.E., Reinecke, M.G., Abdel-Malek, S., Jia, Q., Chow, S.A., 1996b. Proc. Natl. Acad. Sci. 93, 6326–6331.

Rogers, D., Hopfinger, A.J., 1994. J. Chem. Inf. Comput. Sci. 34, 854–866.

Sharma, H., Patil, S., Sanchez, T.W., Neamati, N., Schinazi, R.F., Buolamwini, J.K., 2011. Bioorg. Med. Chem. 19, 2030–2045.

Statsoft, 2001. Statsoft, Inc. STATISTICA, Version 6.0.

Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., Fiore, F., Gardelli, C., Gonzalez Paz, O., Hazuda, D.J., Jones, P., Kinzel, O., Laufer, R., Monteagudo, E., Muraglia, E., Nizi, E., Orvieto, F., Pace, P., Pescatore, G., Scarpelli, R., Stillmock, K., Witmer, M.V., Rowley, M., 2008. J. Med. Chem. 51, 5843–5855.

Todeschini, R., Consonni, V., 2008. E–H Handbook of Molecular Descriptors. Wiley-VCH Verlag GmbH, pp. 124–226, http://dx.doi.org/10.1002/9783527613106.ch1b.

Please cite this article in press as: Srivastav, V.K., Tiwari, M. QSAR and docking studies of coumarin derivatives as potent HIV-1 integrase inhibitors. Arabian Journal of Chemistry (2013), http://dx.doi.org/10.1016/j.arabjc.2013.01.015

Tropsha, A., Gramatica, P., Gombar, V.K., 2003. Mol. Inform. 22, 69–77.

UNAIDS, 2011. World AIDS Day Report 2011, accessed on 02/05/2012, http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2011/JC2216_WorldAIDSday_report_2011_en.pdf.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., 2003. Proteins. 52, 609–623.

Wold, S., Eriksson, L., Clementi, S., 2008. Statistical Validation of QSAR Results. Chemometric Methods in Molecular Design. Wiley-VCH Verlag GmbH, pp. 309–338, http://dx.doi.org/10.1002/9783527615452.ch5.

Yuan, H., Parrill, A.L., 2002. Bioorg. Med. Chem. 10, 4169–4183.

Zhao, H., Neamati, N., Hong, H., Mazumder, A., Wang, S., Sunder, S., Milne, G.W.A., Pommier, Y., Burke, T.R., 1997. J. Med. Chem. 40, 242–249.