Scientific paper

QSAR Analysis of Chicoric Acid Derivatives as HIV–1 Integrase Inhibitors

Kamlesh K. Sahu,[a] V. Ravichandran,[b] Prateek K. Jain,[b] Simant Sharma,[b] V. K. Mourya[c] and R. K. Agrawal[b*]

[a] *Department of Applied Chemistry, Tohoku University Graduate School of Engineering Aoba–yama, Sendai, JAPAN.*

[b] *Department of Pharmaceutical Sciences, Dr. H. S. Gour University, Sagar (M.P.), 470 003 India.*

[c] *Govt. College of Pharmacy, Osmanpura, Aurangabad, Maharashtra, India.*

** Corresponding author: Phone No. 07582233934, E–mail: dragrawal2001@yahoo.co.in*

Received: 21–06–2007

Abstract

HIV-1 Integrase is a potential target for anti- HIV therapy. It is an essential enzyme required for replication of the AIDS-virus. Chicoric acid derivatives act against HIV integrase and thus have the potential to become a part of anti-HIV drug regime. Chicoric acid derivatives have all the features required for it to act as good anti–HIV agent like poly aromatic rings and a central linker. In present study, we have carried out QSAR study of Chicoric acid derivatives using the software WIN CAChe 6.1 and STATISTICA in order to improve its activity. Multiple linear regression analysis was performed to derive QSAR models which were further evaluated for statistical significance and predictive power by internal and external validation. The predictive ability of the selected model was also confirmed by leave 25% out cross validation. The QSAR model indicates that heat of formation, partition coefficient, lowest unoccupied molecular orbital, solvent accessible surface area and shape index play an important role for the HIV integrase inhibitory activities. The results of the present study may be useful on the designing of more potent chicoric acid analogues as HIV integrase inhibitory agents.

Keywords: Human immunodeficiency virus, integrase; inhibition; quantitative structure activity relationship; chicoric acid.

1. Introduction

Human immunodeficiency virus type1 (HIV–1) Integrase is an enzyme required for viral replication.[1] HIV Integrase catalyzes integration of viral DNA into host genome I two separate but chemically similar reactions known as 3'processing and DNA strand transfer.[2,3] In 3' processing IN removes a dinucleotide next to conserved cytosine–adenine sequence from each 3'– end of the viral DNA. IN then attaches the processed 3'– end of the viral DNA to the host cell DNA in the strand transfer reaction. As thee is no known human counterpart of HIV Integrase, IN is an attractive target for anti–retroviral drug design.[4] A large number of HIV IN inhibitors have been discovered.[5] However the mechanism of action is incompletely understood.[6]

Several families of IN inhibitors have been identified. Most of them can be classified into three groups: DNA ligands, C–terminal domain ligands, and compounds that interfere with catalytic domain of the protein. The first family contains nonspecific intercalating agents as well as more specific oligonucleotide targeting Integrase binding sites on both long terminal repeats (LTRs).[7–9] While many Integrase inhibitors have now been developed, only a handful displayed the anti–viral activity in cell culture. This group comprises lignanolides,[10] curcumin,[11] aurintricarboxylic acids,[12] dicaffeoyl quinic acids and analogues,[13,14] diarylsulfones,[15] and finally G–rich oligonucleotides.[16]

Computational chemistry has developed into an important contributor to rational drug design. Quantitative structure activity relationship (QSAR) modeling results in a quantitative correlation between chemical structure and biological activity.[17] Free energy simulations are of particular interest for the interpretation of macroscopic data in terms of microscopic interactions.[18] This can be done by

expressing calculated free energies as a sum of components that correspond to the contributions of different energy terms or different parts of the system. The partitioning of the free energy into additive contributions originating from different groups of atoms or force field terms has the potential to provide relationship between structure and biological activity of molecules. Bren *et al*. formulated the theoretical foundation for the free energy decomposition in the free energy perturbation (FEP) methodology using Thiele cumulants, a powerful tool from the arsenal of probability theory and mathematical statistics.[19]

Free energy perturbation (FEP) calculations using the Amber 95 force field and the TIP3P water model were carried out to evaluate the solvation free energy of deoxyribonucleoside triphosphates in aqueous solution by Bren *et al*.[20] Molecular dynamics simulations was horoughly reviewed by van Gunsteren *et al*.[21] and Hansson *et al*.[22] And a new parametrization of the Langevin dipole (LD) model is developed for ab initio calculations of chemical processes in aqueous solution by Florian *et al*.[23]

QSAR studies have provided valuable insight in the design and development of HIV–1 reverse transcriptase inhibitors,[24–26] HIV–1 protease inhibitors,[27–31] HIV–1 integrase inhibitors,[32–34] and gp120 envelope glycoprotein inhibitors.[35]

As a part of ongoing efforts to design novel molecules with potent anti–HIV activity, a QSAR analysis was performed to relate HIV integrase inhibitory activity of chicoric acid derivatives[36] to its physicochemical properties using Win CAChe version 6.1 (Product of Fujitsu private limited, Japan, http://www.cachesoftware.com/contacts/japan.shtml) modeling software and the QSAR models were generated by STATISTICA version 6 (Soft stat) software. There is high structural diversity and a sufficient range of the biological activity in the selected series of chicoric acid derivatives (Table 1). It insists as to select these series of compounds for our QSAR studies. All the HIV integrase inhibitory activities used in the present study were expressed as pIC_{50} or $-\log IC_{50}$. Where IC_{50} is the micro molar concentration of the compounds producing 50% reduction in the integrase growth is stated as the means of at least two experiments.

2. Results and Discussion

In the present study authors tried to develop best QSAR model to explain the correlation between the physicochemical parameters and HIV integrase inhibitory activity of chicoric acid derivatives. After regression analysis on the software STATISTICA, the best equation received for 3' processing inhibitory activity was –

$$\text{Log } (1/IC_{50}) = 16.124\ (\pm 4.075) - 0.026\ (\pm 0.005)\ \text{HF} + 15.529\ (\pm 3.482)\ \text{LUMO} - 0.764\ (\pm 0.243)\ \text{logP} - 0.326\ (\pm 0.066)\ \text{BKO1} \quad (1)$$

$n = 22$, $r = 0.790$, $r^2 = 0.625$, SEE = 0.667, $F = 7.09$, $P < 0.001$, $q^2 = 0.409$, SDEP = 0.777, $S_{PRESS} = 0.865$, PRESS = 12.71

$$\text{Log } (1/IC_{50}) = 15.613\ (\pm 3.899) - 0.022\ (\pm 0.004)\ \text{HF} + 14.414\ (\pm 3.235)\ \text{LUMO} - 0.540\ (\pm 0.211)\ \text{logP} - 0.020\ (\pm 0.004)\ \text{SAS} \quad (2)$$

$n = 22$, $r = 0.798$, $r^2 = 0.636$, SEE = 0.657, $F = 7.449$, $P < 0.001$, $q^2 = 0.421$, SDEP = 0.770, $S_{PRESS} = 0.857$, PRESS = 12.48

Where IC_{50} is the molar concentration of the drug leading to 50% inhibition of enzyme Integrase, HF = Heat of formation, LUMO = Lowest unoccupied molecular orbital, logP = Partition coefficient, BKO1 = Basic Kappa Order 1 (Shape index), SAS = Solvent accessible surface area. In the above equations n is the number of compounds used to derive the model and values in parentheses are the 95% confidence limit of respective coefficient, r = correlation coefficient, SEE = Standard error of estimation, F = F–ratio between variances of calculated and observed value, r^2 squared correlation coefficient. We extended our study for four parametric correlations as they are permitted for a data set of 22 compounds in accordance with the lower limit of rule of thumb. The calculated and predicted (LOO) activities of the compounds by the above models are shown in Table 2.

Both the models have three outlier's compounds 2, 10 and 18, because their residual values exceeded twice the standard error of estimate. When these outliers have been removed from the data set, we have got two highly significant equations 3 and 4.

$$\text{Log } (1/IC_{50}) = 15.844\ (\pm 2.716) - 0.031\ (\pm 0.004)\ \text{HF} + 15.894\ (\pm 2.355)\ \text{LUMO} - 0.761\ (\pm 0.163)\ \text{logP} - 0.355\ (\pm 0.046)\ \text{BKO1} \quad (3)$$

$n = 19$, $r = 0.915$, $r^2 = 0.837$, SEE = 0.426, $F = 17.94$, $P < 0.001$, $q^2 = 0.750$, SDEP = 0.468, $S_{PRESS} = 0.531$, PRESS = 3.949

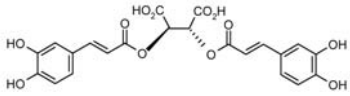
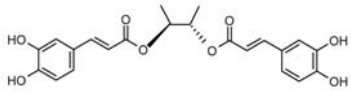
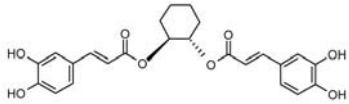
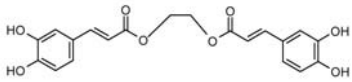
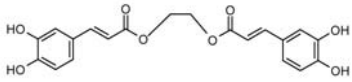
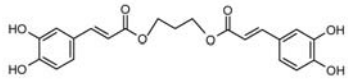
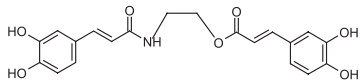
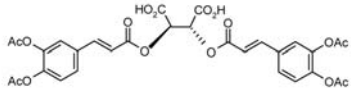
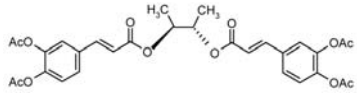
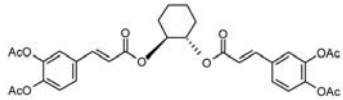
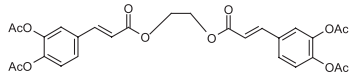
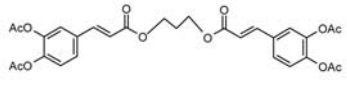
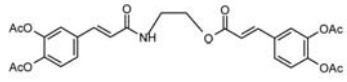
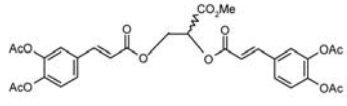
$$\text{Log } (1/IC_{50}) = 15.383\ (\pm 2.470) - 0.026\ (\pm 0.003)\ \text{HF} + 14.782\ (\pm 2.060)\ \text{LUMO} - 0.528\ (\pm 0.135)\ \text{logP} - 0.021\ (\pm 0.003)\ \text{SAS} \quad (4)$$

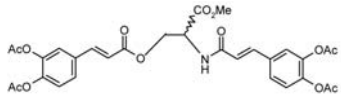
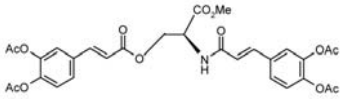
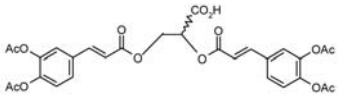
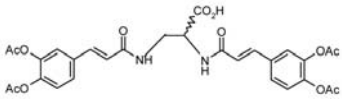
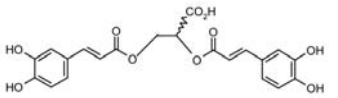
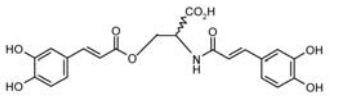
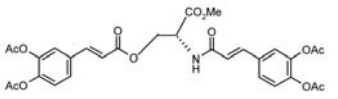
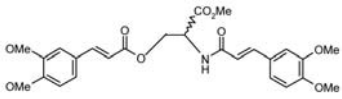
$n = 19$, $r = 0.925$, $r^2 = 0.857$, SEE = 0.399, $F = 20.96$, $P < 0.001$, $q^2 = 0.781$, SDEP = 0.438, $S_{PRESS} = 0.497$, PRESS = 3.461

The best equation received for integration inhibitory activity was

$$\text{Log } (1/IC_{50}) = 19.716\ (\pm 3.246) - 0.025\ (\pm 0.004)\ \text{HF} + 17.814\ (\pm 2.737)\ \text{LUMO} - 0.882\ (\pm 0.192)\ \text{logP} - 0.336\ (\pm 0.051)\ \text{BKO1} \quad (5)$$

Table 1. Structures and physicochemical parameters of chicoric acid derivatives

Compd. No	Structure	LUMO	HF	logP	BKO1	SAS
1		–1.037	–421.599	2.528	30.225	449.253
2		–1.034	–567.23	1.661	42.168	612.623
3		–0.94	–408.778	3.013	38.184	590.738
4		–0.939	–410.569	3.447	38.448	605.03
5		–0.95	–399.288	2.187	36.193	562.718
6		–0.941	–371.188	2.239	37.188	580.143
7		–0.947	–358.774	1.537	36.193	564.547
8		–0.846	–264.663	3.88	26.254	425.252
9		–0.841	–269.341	3.969	26.602	448.034
10		–0.871	–253.876	3.054	24.271	401.569
11		–0.85	–259.92	3.106	25.262	420.297
12		–0.872	–213.221	2.404	24.271	403.363
13		–0.985	–476.939	1.956	40.176	611.533
14		–0.944	–438.923	1.306	40.176	613.818

Compd. No	Structure	LUMO	HF	logP	BKO1	SAS
15		–0.928	–400.01	0.656	40.176	582.256
16		–0.811	–225.127	1.649	32.213	520.868
17		–0.991	–484.895	1.924	39.18	591.719
18		–0.987	–405.878	0.624	39.18	587.064
19		–0.923	–338.936	2.791	27.246	430.328
20		–0.891	–300.716	2.141	27.246	432.988
21		–1.184	–327.25	0.013	27.034	414.611
22		–1.186	–259.901	0.389	20.045	314.785

n = 20, r = 0.880, r^2 = 0.774, SEE = 0.494,
F = 12.83, P < 0.001, q^2 = 0.620, SDEP = 0.582,
S_{PRESS} = 0.634, PRESS = 6.441

Model – 3 shows good correlation coefficient (r) of 0.915 between descriptors (HF, logP, LUMO and BKO1) and HIV integrase 3' processing inhibitory activity. Squared correlation coefficient (r^2) of 0.837 explains 83.7% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 17.94. Cross validated squared correlation coefficient of this model was 0.750, which shows the good internal prediction power of this model.

Model – 4 shows good correlation coefficient (r) of 0.925 between descriptors (HF, logP, LUMO and SAS) and HIV integrase integration inhibitory activity. Squared correlation coefficient (r^2) of 0.857 explains 85.7% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 20.96. Cross validated squared correlation coefficient of this model was 0.781, which shows the good internal prediction power of this model. Model – 5 shows good correlation coefficient (r) of 0.880 between descriptors (HF, logP, LUMO and BKO1) and HIV integrase integration inhibitory activity. Squared correlation coefficient (r^2) of 0.774 explains 77.4% variance in biological activity. This model also indicates statistical significance >99.9% with F values F = 12.83. Cross validated squared correlation coefficient of this model was 0.620, which shows the good internal prediction power of this model.

The predictive ability of model – 1, 2 and 5 was also confirmed by external validation (model – 6, 7 and 8 respectively) (Table 3). The r^2CVext value of the selected model is greater than the prescribed value (r^2CVext >0.5).

The QSAR model for training set of 3' processing inhibition activity using model –1 & 2:

$$\text{Log } (1/IC_{50}) = 16.026\ (\pm\ 3.475) - 0.028\ (\pm\ 0.005)\ \text{HF} + 15.933\ (\pm\ 3.099)\ \text{LUMO} - 0.862\ (\pm\ 0.196)\ \text{logP} - 0.326\ (\pm\ 0.060)\ \text{BKO1} \quad (6)$$

n = 18, r = 0.854, r^2 = 0.729, SEE = 0.520,
F = 8.72, r^2CVext = 0.729

Log ($1/IC_{50}$) = 15.874 (± 3.201) –0.024 (± 0.004) HF + 15.202 (± 2.787) LUMO –0.642 (± 0.164) logP –0.020 (± 0.003) SAS (7)

n = 18, r = 0.871, r^2 = 0.759, SEE = 0.489,
F = 10.28, r^2CVext = 0.709

The QSAR model for training set of integration inhibition activity

Log ($1/IC_{50}$) = 20.088 (± 3.476) –0.029 (± 0.005) HF + 18.544 (± 2.909) LUMO –0.919 (± 0.187) logP –0.371(± 0.063) BKO1 (8)

n = 16, r = 0.894, r^2 = 0.799, SEE = 0.443,
r^2CVext = 0.608

The robustness of the selected model (2 and 5) was checked by Y – randomization test (Table 4). The low r^2 and q^2 values indicate (data not shown) that the good results in our original model are not due to a chance correlation or structural dependency of the training set.

The predictive ability of these models (1, 2 &5) was also confirmed by leave 25% out cross validation. All the three models showed good predictivity.

The equations received from the leave 25% out cross validation technique are Leave 25% out crossvalidation equation for model – 1 (number of cycles 4)

Log ($1/IC_{50}$) = 17.027 (± 4.835) –0.027 (± 0.023) HF + 16.254 (± 4.117) LUMO –0.814 (± 0.288) logP –0.335 (± 0.077) BKO1 (9)

r^2CVext = 0.619

Leave 25% out crossvalidation equation for model – 2 (number of cycles 4)

Log ($1/IC_{50}$) = 16.336 (± 4.551) –0.022 (± 0.005) HF + 14.973 (± 3.762) LUMO –0.578 (± 0.246) logP –0.019 (± 0.004) SAS (10)

r^2CVext = 0.678

Leave 25% out crossvalidation equation for model – 5 (number of cycles 4)

Table 2. Observed, calculated and predicted (LOO) 3' processing and integration inhibitory activity of chicoric acid derivatives

Cd	Obs[a]	Obs[b]	Model – 1		Model – 2		Model – 3		Model – 4		Model – 5	
No			Cal.	Pred.	Cal.	Pred	Cal.	Pred.	Cal.	Pred	Cal.	Pred.
1	–0.04	0.100	–0.636	–0.807	–0.393	–0.517	0.389	0.529	–0.113	–0.151	–0.625	–0.911
2	–0.99	–0.79	0.023	0.607	0.064	0.686	***	***	***	***	–0.187	0.252
3	–2.52	–2.52	–2.437	–2.409	–2.309	–2.248	–2.437	–2.409	–2.297	–2.234	–2.323	–2.258
4	–2.52	–2.52	–2.792	–2.988	–2.770	–2.942	–2.791	–2.988	–2.773	–2.947	–2.732	–2.884
5	–2.00	–2.00	–1.562	–1.533	–1.662	–1.637	–1.551	–1.515	–1.652	–1.622	–1.339	–1.294
6	–2.52	–2.52	–1.629	–1.556	–1.773	–1.702	–1.618	–1.532	–1.766	–1.686	–2.262	–2.205
7	–2.52	–2.00	–2.089	–2.035	–2.180	–2.132	–2.249	–2.209	–2.341	–2.312	–1.724	–1.679
8	–1.58	–0.89	–1.554	–1.548	–1.293	–1.224	–1.778	–1.838	–1.509	–1.487	–0.999	–1.026
9	–1.39	–0.88	–1.535	–1.573	–1.615	–1.674	–1.746	–1.860	–1.851	–1.999	–0.989	–1.017
10	0.40	0.4	–0.949	–1.200	–0.976	–1.226	***	***	***	***	–0.318	–0.486
11	–1.44	–0.88	–0.826	–0.705	–0.938	–0.848	–1.045	–0.948	–1.175	–1.116	–0.173	0.013
12	–1.44	–0.81	–1.542	–1.563	–1.554	–1.577	–1.939	–2.069	–1.963	–2.098	–0.778	–0.768
13	–1.00	–1.00	–1.177	–1.202	–1.320	–1.366	–0.968	–0.962	–1.104	–1.127	–1.159	–1.183
14	–0.40	–0.51	–1.048	–1.159	–1.245	–1.389	–0.986	–1.138	–1.178	–1.372	–0.805	–0.866
15	–1.04	–1.00	–1.331	–1.434	–0.886	–0.827	–1.429	–1.614	–0.918	–0.855	–0.918	–0.875
16	–3.00	---	–2.296	–1.842	–2.314	–1.869	–2.852	–2.721	–2.877	–2.767	---	---
17	–0.32	–0.45	–0.711	–0.777	–0.828	–0.910	–0.442	–0.479	–0.547	–0.611	–0.705	–0.753
18	–0.61	–0.61	–1.743	–1.991	–1.686	–1.916	***	***	***	***	–1.459	–1.696
19	–0.62	–0.23	–0.279	–0.203	–0.308	–0.242	–0.251	–0.140	–0.278	–0.182	0.111	0.203
20	–0.52	---	–0.295	–0.235	–0.375	–0.339	–0.418	–0.385	–0.499	–0.492	---	---
21	–2.52	–2.52	–2.428	–2.361	–2.500	–2.485	–2.547	–2.569	–2.619	–2.697	–2.285	–2.114
22	–2.52	–2.05	–2.265	–1.969	–2.244	–1.925	–2.464	–2.398	–2.445	–2.355	–2.006	–1.955

Where *** is outliers, --- not included in model development, Obs[a] – observed 3' processing inhibitory activity, Obs[b] – observed integration inhibitory activity in μM concentration. Model – 1 & 2 are the QSAR model for the 3' processing inhibitory activity before outlier removal, Model – 3 & 4 are the QSAR model for the 3' processing inhibitory activity after outlier (compound number 2, 10, 18) removal, and Model –5 is the QSAR model for the integration inhibitory activity of chicoric acid (compound 16 and 20 are not included in model development because they are not having specified activity range).

$$\text{Log } (1/IC_{50}) = 19.597\ (\pm 4.320) - 0.020\ (\pm 0.004)\ HF + 16.865\ (\pm 3.548)\ LUMO - 0.682\ (\pm 0.233)\ logP - 0.361\ (\pm 0.0634)\ BKO1 \quad (11)$$

$r^2CVext = 0.614$

Consequently equation–4 can be considered as most suitable model for 3' processing inhibitory activity with both high statistical significant and excellent predictive ability. So we have taken the original model of the model – 4, that is model – 2 is the best model. Model – 5 was selected as best model for integration inhibition activity. The variables used in the selected models have no mutual correlation (Table 5). These models showed good correlation coefficient between descriptors and HIV integrase 3' processing and integration inhibitory activity.

In model – 2, the negative contribution of HF, SAS and logP on the biological activity showed that the increase in the values of these parameters lead to better HIV integrase 3' processing inhibitory chicoric acid compounds. The negative contribution of SAS indicates that the steric bulkiness will be detrimental to HIV integrase activity of chicoric acid. The positive coefficient of LUMO showed that the substitution with groups having high electro negativity is conducive for the inhibitory activity of chicoric acid. The less electro negative groups are detrimental to biological activity. In model –5, all the three parameters except SAS showed the same type of contribution to HIV integrase intergration inhibition activity of chicoric acid. The other variable in the model – 5 is BKO1, which shown negative contribution to biological activity. In both the model logP is contributed negatively; because, our opinion is that the logP values of the selected series of compounds are below the optimum logP value of HIV integrase inhibitors so it's contributing negatively. Thus we conclude that if the groups that bring about the above mentioned changes in the molecule, are attached to it, the biological activity will be increased. Based on the developed QSAR model, new HIV integrase inhibitors of chicoric acid derivatives can be designed with caution.

3. Experimental

3. 1. General Procedure

Win CAChe 6.1 (molecular modeling software, a product of Fujitsu private limited, Japan), STATISTICA version 6 (Stat Soft, Inc., Tulsa, USA).

3. 2. Optimization of Molecules Structure

A data set of 22 compounds of chicoric acid for HIV integrase activity (Table 1) was used for the present QSAR study.[30] The molar concentrations of the chicoric acid compounds required to produce 50% reduction in the HIV integrase enzyme growth is stated as the means of at least two experiments were converted to free energy related negative logarithmic values for undertaking the QSAR study. All 22 compounds structure were built on workspace of Win CAChe 6.1 (molecular modeling software, a product of Fujitsu private limited, Japan) and energy minimization of the molecules was done using Allinger's MM2 force field followed by semi empirical PM3 method available in MOPAC module until the root mean square gradient value becomes smaller than 0.001 kcal/mol Å. Most stable structure for each compound was generated and used for calculating various physico–chemical descriptors like thermodynamic, steric and electronic values of descriptors.

3. 3. Descriptors Calculation, QSAR Models Development and Validation

All the calculated descriptors (18 descriptors calculated by Win CAChe 6.1, the complete descriptors data set of all compounds will be provided on request) were considered as independent variable and biological activity as dependent variable. STATISTICA version 6 (Stat Soft, Inc., Tulsa, USA) software was used to generate QSAR models by multiple linear regression analysis. Statistical measures used were n–number of compounds in regression, r–correlation coefficient, r^2–squared correlation coefficient, F– test (Fischer's value) for statistical significance, SEE– standard error of estimation, q^2 or r^2_{CV} – cross validated correlation coefficient and correlation matrix to show correlation among the parameters.

The squared correlation coefficient (or coefficient of multiple determination) r^2 is a relative measure of fit by the regression equation. Correspondingly, it represents the part of the variation in the observed data that is explained by the regression. The correlation coefficient values closer to 1.0 represent the better fit of the regression. The F–test reflects the ratio of the variance explained by the model and the variance due to the error in the regression. High values of the F–test indicate that the model is statistically significant. Standard deviation is measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus standard deviation is an absolute measure of quality of fit and should have a low value for the regression to be significant.

The predictive ability of the generated correlations was evaluated by cross validation method employing a 'leave–one–out' scheme.

Validation parameters considered were cross validated r^2 or q^2, standard deviation based on predicted residual sum of squares (S_{PRESS}) and standard error of prediction (SDEP). The predictive ability of the selected model was also confirmed by external r^2CVext.

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

Table 3. Experimental and predicted activities of training and test set of compounds.

Compd. No.	Obs. Act.		Predicted activity		
	3' processing	Integration	Model – 1	Model – 2	Model – 5
1	–0.04	0.100	–0.558	–0.281	–0.312
2[a,b]	–0.99	–0.79	0.465	0.580	0.384
3	–2.52	–2.52	–2.393	–2.276	–2.289
4	–2.52	–2.52	–2.787	–2.784	–2.715
5	–2.00	–2.00	–1.461	–1.565	–1.255
6[b]	–2.52	–2.52	–1.517	–1.666	–2.328
7	–2.52	–2.00	–2.004	–2.126	–1.790
8	–1.58	–0.89	–1.846	–1.586	–1.146
9	–1.39	–0.88	–1.824	–1.911	–1.127
10[a,b]	0.40	0.4	–0.119	–0.122	–0.429
11	–1.44	–0.88	–1.053	–1.165	–0.279
12	–1.44	–0.81	–1.802	–1.845	–1.044
13	–1.00	–1.00	–0.912	–1.039	–0.890
14	–0.40	–0.51	–0.779	–0.971	–0.648
15	–1.04	–1.00	–1.069	–0.623	–0.895
16[a]	–3.00	——	–2.438	–2.509	——
17	–0.32	–0.45	–0.429	–0.518	–0.369
18[b]	–0.61	–0.61	–1.491	–1.453	0.241
19	–0.62	–0.23	–0.346	–0.349	–1.418
20	–0.52	——	–0.363	–0.431	——
21[a]	–2.52	–2.52	–2.358	–2.487	–2.285
22	–2.52	–2.05	–2.362	–2.406	–2.075

—— indicates compound not included in model development, [a] indicates test set compounds for 3' processing inhibition and [b] indicates test set compounds for integration inhibition activity.

Table 4. Results of Y – randomization test

Iteration	Model – 2		Model – 5	
	r^2	q^2	r^2	q^2
1	0.124	0.005	0.113	0.008
2	0.085	0.008	0.068	0.005
3	0.065	0.000	0.056	0.001
4	0.139	0.023	0.075	0.001
5	0.143	0.051	0.120	0.012

Table 5. Inter correlation of molecular descriptors used in models

Descriptors	HF	Log P	LUMO	BKO1	SAS
HF	1				
Log P	0.2191	1			
LUMO	0.376	0.671	1		
BKO1	–0.644	–0.210	–0.042	1	
SAS	–0.433	–0.110	0.071	0.986	1

The robustness of a QSAR model was checked by Y – randomization test. In this technique, new QSAR models were developed by shuffling the dependent variable vector randomly and keeping the original independent variable as such. The new QSAR models are expected to have low r^2 and q^2 values. If the opposite happens then an acceptable QSAR model can not be obtained for the specific modeling method and data. The predictivity of the model is also confirmed by leave 25% out crossvalidation method. In this method 25% of the compounds are removed at once, the model is developed by using the remaining compounds and activity of the removed compounds will be predicted by developed model.

4. References

1. E. De Clercq, *J. Med. Chem.* **1995**, *38*, 2491–2517.
2. R. Craigie, T. Fujiwara, F. Bushman, *Cell.* **1990**, *62*, 829–837.
3. R. A. Katz, G. Merkel, T. Kulkosk, J. Leis, A. M. Salka, *Cell.* **1990**, *63*, 87–95.
4. I. J. Chen, N. Neamati, M. C. Nicklaus, A. Orr. L. Anderson, J. J. Jr. Barchi, J. A. Kelly, Y. Pommier, A. D. Jr. Mac–Kerell, *Bioorg. Med. Chem.* **2000**, *8*, 2385–2398.
5. N. Neamati, *Expert Opin. Ther. Patents*, **2002**, *12*, 709–713.
6. A. L. Parril, *Curr. Med . Chem.* **2003**, *10*, 1811–1824.
7. S. Carteau, J. F. Mouscadet, H. Goulaouic, F. Subra, C. Auclair, *Biochem. Biophys. Res. Commun.* **1993**, *192*, 1409–1414.
8. M. R. Fesen, K. W. Kohn, F. Leteurtre, Y. Pommier, *Proc. Natl. Acad. Sci.* U. S. A. **1993**, *90*, 2399–2403.

9. J. F. Mouscadet, S. Carteau, H. Goulaouic, F. Subra, C. Auclair, *J. Biol. Chem.* **1994**, *269*, 21635–21638.
10. E. Eich, H. Pertz, M. Kaloga, J. Schulz, M. R. Fesen, A. Majumder, Y. Pommier, *J. Med. Chem.* **1996**, *39*, 86–95.
11. A. Majumder, K. Raghawan, J. Weinstein, K. W. Kohn, Y. Pommier, *Biochem. Pharmacol.* **1995**, *49*, 1165–1170.
12. M. Cushman, P. Sherman, *Biophys. Res. Commun.* **1992**, *185*, 58–63.
13. W. E. Jr. Robinson, M. Cordeiro, S. Abdel–Malek, Q. Jia, S. A. Chow, M. G. Reineck, W. M. Mitchell, *Mol. Pharmacol.* **1996**, *50*, 846–855.
14. W. E. Jr. Robinson, M. Cordeiro, S. Abdel–Malek, Q. Jia, S. A. Chow, *Proc. Natl. Acad. Sci*. U. S. A. **1996**, *93*, 6326–6331.
15. A. Mazumder, N. Neamati, J. P. Sommadossi, G. Gosseli, R. F. Schinazi, Y. Pommier, *Mol. Pharmacol.* **1996**, *49*, 621–628.
16. C. Hansch, *Acc. Chem. Res*. **1969**, *2*, 232–236.
17. A. L. Paril, *Curr. Med. Chem.* **2003**, *10*, 1811–1185.
18. S. Boresch, M. Karplus, *J. Mol. Biol*. **1995**, *254*, 801–807.
19. M. Bren, J. Florian, J. Mavri, U. Bren, *Theor. Chem. Acc*. **2007**, *117*, 535–540.
20. U. Bren, V. Martinek, J. Florian, *J. Phys. Chem. B*. **2006**, *110*, 12782–12788.
21. W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992–1023.
22. T. Hansson, C. Oostenbrink, W. F. van Gunsteren, *Curr. Opin. Struct. Biol*. **2002,** *12*, 190–196.
23. J. Florian, A. Warshel, *J. Phys. Chem. B.* **1997**, *101*, 5583–5595.
24. V. Ravichandran, R. K. Agrawal, *Bioorg. Med. Chem. Lett*. **2007**, *17*, 2197–2202.
25. P. Pungpo, S. Hannongbua, *J. Mol. Graphics & Modell*. **2000**, 18, 581–590.
26. M. L. Barreca, A. Carotti, A. Chimirri, A. M. Monforte, *Bioorg. Med. Chem.*, **1999**, *7*, 2283–2292.
27. A. C. Nair, P. Jayatilleke, X. Wang, S. Miertus, W. J. Welsh, *J. Med. Chem*. **2002**, *45*, 973–983.
28. P. R. N. Jayatilleke, A. C. Nair, R. Zauhar, W. J. Welsh, *J. Med. Chem.* **2000**, *43*, 4446–4451.
29. A. Hilgeroth, R. Fleischer, M. Wiese, F. W. Heinemann, *J. Computer–Aided Mol. Design.* **1999**, *13*, 233–242.
30. A. K. Debnath, *J. Med. Chem*. **1999**, *42*, 249–259.
31. T. I. Oprea, C. L. Waller, G. R. Marshall, *Drug Design Discovery.* **1994**, *12*, 29–51.
32. J. K. Buolamwini, H. Assefa, *J. Med. Chem.* **2002**, *45*, 841–852.
33. K. Raghavan, J. K. Buolamwini, M. R. Fesen, Y. Pommier, K. W. Kohn, *J. Med. Chem*. **1995**, *38*, 890–897.
34. K. K. Sahu, V. Ravichandran, V. K. Mourya, R. K. Agrawal, *Med. Chem. Res*. **2007** (In press).
35. A. K. Debnath, S. Jiang, N. Strick, K. Lin, P. Haberfield, *J. Med. Chem*. **1994**, *37*, 1099–1108.
36. Y. W. Xu, G. S. Zhao, C. G. Shin, H. C. Zang, C. K. Lee, Y. S. Lee, *Bioorg. Med. Chem.* **2003**, *11*, 3589–3593.

Povzetek

HIV-1 integraza je potencialni kandidat za anti-HIV terapijo. Je encim potreben za replikacijo AIDS virusa. Derivati kikorne kisline zavirajo HIV integrazo, zato lahko postanejo potencialno del anti-HIV terapije. Derivati kikorne imajo vse potrebne strukturne lastnosti (poliaromatski obroči ter centralno vezalno mesto), da bi lahko potencialno služili kot anti-HIV reagent. V raziskavi smo za študij povezav med kemijsko strukturo in biološko aktivnostjo (QSAR) derivatov kikorne kisline uporabili programsko opremo WIN CAChe 6.1 in STATISTICA. Za izdelavo QSAR modelov smo uporabili multiplo linearno regresijo. Modele smo testirali glede njihove signifikantnosti in napovedne sposobnosti z uporabo internih in eksternih validacijskih postopkov. Najboljši QSAR model za napovedovanje »3' procesiranje« biološke aktivnosti je imel naslednje validacijske parametre po odstranitvi izven ležečih točk. Korelacijki koeficient (r) je znašal 0,93, standardna napaka (SEE) je bila 0,40 in kvadrat korelacijskega koeficienta pri navzkrižni validaciji (q^2) je bil 0,78. Najboljši model za napovedovanje integralne inhibitorne aktivnosti je imel naslednje validacijske parametre: r = 0,88, SEE = 0,49 in q^2 = 0,62. Napovedno sposobnost QSAR modelov smo potrdili z navzkrižnim validacijkim testom, pri čemer smo v vsakem koraku izpustili 25 % podatkov. Z analizo QSAR modelov smo ugotovili, da igrajo pomembno vlogo za napovedovanje inhibitornih sposobnosti posameznih spojin naslednji strukturni deskriptorji: formacijska toplota, particijski koeficient (log P), energija najnižje nezasedene orbitale, dostopna površina za topilo in indeks oblike molekule. Rezultati opisane študije bi lahko bili uporabljeni za izdelavo aktivnejšega analoga kikorne kisline.