Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 14 (2004) 6089–6094
Bioorganic & Medicinal Chemistry Letters

# QSAR modelling of HIV-1 reverse transcriptase inhibition by benzoxazinones using a combination of P_VSA and pharmacophore feature descriptors

S. Balaji,[a] C. Karthikeyan,[a] N. S. Hari Narayana Moorthy[b] and Piyush Trivedi[a,*]

[a]*Department of Pharmacy, S.G.S.I.T.S., Indore 452003, Madhya Pradesh, India*
[b]*School of Pharmaceutical Sciences, RGPV, Bhopal 462036, Madhya Pradesh, India*

Received 11 July 2004; revised 17 September 2004; accepted 21 September 2004
Available online 14 October 2004

**Abstract**—In pursuit of better anti-HIV drugs, a quantitative structure–activity relationship analysis using a novel set of 2D descriptors was performed on a series of HIV-1 reverse transcriptase inhibitory benzoxazinones. The QSAR models derived from the above mentioned descriptors were found to be statistically significant and exhibited superior predictive power. The results of the study justify the application of the descriptors for exploring the binding mode of the benzoxazinones to the enzyme.


Acquired immuno deficiency syndrome (AIDS) is a formidable pandemic that is still wreaking havoc worldwide. The catastrophic potential of this virally caused disease may not have been fully realized. The causative moiety of the disease is human immunodeficiency virus (HIV), which is a retrovirus of the lentivirus family.[1] The three viral enzymes; reverse transcriptase, protease and integrase encoded by the *gag* and *gag-pol* genes of HIV play an important role in the virus replication cycle.[2] Among them, viral reverse transcriptase (RT) catalyzes the formation of proviral DNA from viral RNA, the key stage in viral replication. Its central role in viral replication makes RT a prime target for anti-HIV-therapy.[3]

Two main categories of HIV RT inhibitors have been discovered to date. The first category of inhibitors is nucleoside analogues (e.g., AZT, 3TC, ddI, ddC) and the second category of inhibitors is nonnucleoside analogues.[1] Nevirapine, delaviridine and efavirenz are the only nonnucleoside reverse transcriptase inhibitors (NNRTI) that have received regulatory approval with several NNRTIs (MKC442, Troviridine, S-1153/AG1549, PNU142721, ACT and HBY1293/GW420867X) currently undergoing clinical trials. Efavirenz was the first potent anti-HIV drug to be approved by FDA and studies have shown that efavirenz penetrates into the cerebrospinal fluid, a common viral sanctuary.[4] The therapeutic efficacy of the drug is mainly restricted due to the development of viral resistance associated with mutations that includes K103N, L100I and Y188L.[5] In search of effective efavirenz analogues with minimal viral resistance problems, Patel et al.[6] synthesized and evaluated a novel set of benzoxazinones (analogues of efavirenz) for their HIV-1 reverse transcriptase inhibitory activity.

As a part of ongoing efforts to design novel molecules with potent anti-HIV activity, a QSAR analysis was performed to relate HIV-1 reverse transcriptase (HIV-1 RT) inhibitory activity of benzoxazinones to its molecular structure. A novel set of 2D molecular descriptors[7] in the QuaSAR module of molecular modelling package MOE[8] (molecular operating environment) was used to find QSAR model that fits HIV-1 RT inhibition. The QuaSAR descriptors are quite simple to calculate, neither 3D calculation nor alignment step is required for the calculation of the descriptors. The 2D descriptors found in the module are physical properties, atom and bond counts, topological descriptors, pharmacophore feature descriptors and property labelled van der Waals surface area (P_VSA) descriptors. P_VSA descriptors

*Keywords*: QSAR; Benzoxazinones; HIV-1 reverse transcriptase inhibitors.
* Corresponding author. Tel./fax: +91 0731 2368582; e-mail: prof_piyushtrivedi@yahoo.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.09.068

are a set of 52 2D descriptors describing electrostatic, lipophilic, steric and pharmacophoric properties in terms of molecular surface.[9,10]

Despite wide applicability of P_VSA descriptors, there are no extensive applications of P_VSA descriptors in QSAR studies and till date, no QSAR modelling of anti-HIV molecules using this class of descriptors has been explored. In the view of above, the current work strives to propose QSAR models for describing the HIV-1 RT inhibition of benzoxazinones using P_VSA descriptors in tandem with the other molecular descriptors found in the QuaSAR module of MOE. The results of the study revealed that modelling of HIV-1 RT inhibition by benzoxazinones can be carried out using combination of P_VSA descriptors and pharmacophore feature descriptors with acceptable degree of precision. The methodology used for the QSAR analysis and the results obtained are discussed herein.

A data set of the 14 compounds out of the 16 compounds reported by Patel et al.[6] for HIV-1 RT inhibition was used for the QSAR study, since compounds **15** and **16** were not exhibiting a well defined HIV-1 RT inhibitory activity (Table 1). The molar concentrations of the benzoxazinones required to inhibit HIV-1 reverse transcriptase enzyme by 50% were converted to free energy related negative logarithmic values for undertaking the QSAR study.

The 2D descriptors were calculated for the built structures of the compounds in the series using the QuaSAR module of MOE on a Pentium IV workstation. The pharmacophore feature descriptors consider only the heavy atoms of a molecule and assign a type to each atom with their bonded hydrogen suppressed during the calculation. The feature set includes donor, acceptor, polar (both donor and acceptor), positive (base), negative (acid), hydrophobes and others.[7]

The P_VSA descriptors include a novel set of 32 descriptors named 'widely applicable set of descriptors' derived by summing the approximate exposed surface area for each according to the classification based upon log P, molar refractivity and partial charge.[9]

P_VSA descriptors were based on approximate van der Waals surface area (VSA) calculation using connection table approximation for an atom $V$, along with some atomic property $P$. Each descriptor in the series is defined as the atomic VSA contributions of each atom i with $P_i$ in the range (u,v). Thus $\mathrm{P\_VSA}_{(u,v)}$ can be defined as given in Eq. 1.

$$\mathrm{P\_VSA}_{(u,v)} = \sum V_i \delta(P_i \in (u,v)) \quad (1)$$

where $V_i$ = atomic contribution of atom i to the VSA of the molecule.

A set of $n$ descriptors associated with the property $P_i$ is calculated so that they account for all the value of $P_i$ in any molecule. It is defined as follows (Eq. 2).

$$\mathrm{P\_VSA}_k = \sum V_i \delta(P_i \in [a_{k-1}, a_k]) \quad k = 1, 2, \ldots, n \quad (2)$$

where $a_0 < a_k < a_n$ are interval boundaries such that $(a_0, a_n)$ bound are values of $P_i$ in any molecule. Each VSA type descriptor can be characterized as the amount of surface area with $P$ in a certain range. In a given set of descriptors, the interval ranges span all values; the sum of the descriptors will be the VSA of the molecule.[7,9]

Thus a set of 32 descriptors were constructed based on this methodology as follows:

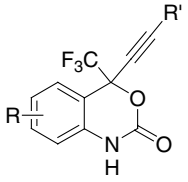
Slog P_VSA$_k$ descriptors intend to capture hydrophobic and hydrophilic effects either in the receptor or on the way to the receptor and the atomic contributions to log P was calculated by the Wildman and Crippen Slog P model.[11]

SMR_VSA$_k$ descriptors intend to capture polarizability and the atomic contributions were calculated using the Wildman and Crippen SMR model.[11]

PEOE_VSA$_k$ descriptors intend to capture direct electrostatic interactions calculated by the Partial equalization of orbital electronegativities (PEOE) method.[12]

Indicator variables were employed to account for the variation in the position R$^1$. In the present case, the indicator variables Inp, Iip, Icp, Ime, Iet, take the value of unity in the presence of *n*-propyl, isopropyl, cyclopropyl, methyl, ethyl, respectively, and shall be valued zero in the absence of the corresponding substituents.

**Table 1.** Structural variations in the benzoxazinones backbone and their HIV-1 RT inhibitory activity




| Compound no. | R | R$^1$ | $IC_{50}$ (nM) | $-\log IC_{50}$ (M) |
|---|---|---|---|---|
| **1** | 5-F | Cyclopropyl | 78 | 7.1079 |
| **2** | 5-F | Ethyl | 127 | 6.8961 |
| **3** | 5-F | *n*-Propyl | 156 | 6.8068 |
| **4** | 5-F | Isopropyl | 102 | 6.9913 |
| **5** | 6-$NO_2$ | Cyclopropyl | 209 | 6.6798 |
| **6** | 6-$NO_2$ | Ethyl | 276 | 6.559 |
| **7** | 6-$NO_2$ | *n*-Propyl | 304 | 6.5171 |
| **8** | 6-$NO_2$ | Isopropyl | 199 | 6.7011 |
| **9** | 6-$NH_2$ | Cyclopropyl | 802 | 6.0958 |
| **10** | 6-$NH_2$ | Ethyl | 1894 | 5.7226 |
| **11** | 6-$NH_2$ | *n*-Propyl | 1506 | 5.8221 |
| **12** | 6-$NH_2$ | Isopropyl | 896 | 6.0476 |
| **13** | 6-N(H)$CH_3$ | Cyclopropyl | 608 | 6.216 |
| **14** | 6-N(H)$CH_3$ | Isopropyl | 473 | 6.3251 |
| **15** | 6-N(H)Ac | Cyclopropyl | >2000 | — |
| **16** | 6-N(H)Ac | Isopropyl | >2000 | — |

Statistical analysis of the activity parameters and the calculated descriptors was carried out using in-house program VALSTAT.[13] The program employs stepwise regression method where the independent variables are individually added or deleted from the model at each step of the regression depending on the Fisher ratio values selected to enter and to remove until the 'best' equation is obtained. The statistical significance of the generated equations was adjudged by the statistical parameters like correlation coefficient ($r$), squared correlation coefficient ($r^2$), standard deviation (SD) and $F$-test, the ratio between the variances of the calculated and observed activities.

The squared correlation coefficient (or coefficient of multiple determination), $r^2$, is a relative measure of quality 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.[14] 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.[14] 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.[14,15]

The predictive ability of the generated correlations was evaluated by cross validation method employing a 'leave-one-out' scheme.[16–18] The validation parameters considered were cross validated $r^2$ or $q^2$, standard deviation based on predicted residual sum of squares ($S_{PRESS}$) and standard deviation of error of prediction (SDEP).

The basic structure of benzoxazinones with the corresponding substituents and the activity parameters is given in Table 1. The compounds in the congeneric series exhibit two sites of structural variations, one in the aromatic ring R and another at the alkynyl side chain $R^1$. Indicator variables were assigned for the different substituents at $R^1$ to account for the features that could not be explained by the continuous variables.

Statistical processing by stepwise regression method indicated that mono-parametric and bi-parametric correlations were possible for modelling the HIV-1 RT inhibition by benzoxazinones. The statistically significant correlations are given in Table 2 and the descriptors found in the generated models are presented in Table 3.

The best mono-parametric correlations (models 1 and 2) were obtained with the P_VSA descriptors relating to the hydrophobicity SlogP_VSA0 representing the VSA of the atoms contributing to the logP (o/w) in the range ($\leqslant -0.4$) and PEOE descriptor for the VSA of atoms

**Table 2.** Statistically significant QSAR models for modelling HIV-1 RT inhibition with statistical parameters

| Model no. | Models | $r$ | $r^2$ | SD | $F$-test |
|---|---|---|---|---|---|
| 1 | BA = [7.25257 (±0.247281)]<br>+ SlogP_VSA0[−0.0263189 (±0.00741817)] | 0.9141 | 0.83557 | 0.1837 | 60.9793 |
| 2 | BA = [6.77379 (±0.140589)]<br>+ PEOE_VSA + 3[−0.0504094 (±0.0143854)] | 0.9122 | 0.83213 | 0.1856 | 59.4864 |
| 3 | BA = [5.68947 (±0.6645)]<br>+ SlogP_VSA0[−0.0274342 (±0.00415453)]<br>+ PEOE_VSA_FNEG[2.89757 (±1.2051)] | 0.9769 | 0.95444 | 0.1010 | 115.214 |
| 4 | BA = [5.14091 (±0.633658)]<br>+ PEOE_VSA + 3[−0.0528416 (±0.00762661)]<br>+ PEOE_VSA_FNEG[2.99071 (±1.15264)] | 0.9789 | 0.95843 | 0.0965 | 126.805 |
| 5 | BA = [7.52658 (±0.389664)]<br>+ a_don[−0.744188 (±0.257738)] | 0.8780 | 0.77094 | 0.2168 | 40.3874 |
| 6 | BA = [4.56661 (±1.26537)]<br>+ a_don[−0.638474 (±0.151413)]<br>+ a_hyd[0.181222 (±0.0763049)] | 0.9671 | 0.93539 | 0.1203 | 79.6283 |
| 7 | BA = [−3.47307 (±3.58993)]<br>+ PEOE_RPC−[53.2791 (±19.2361)] | 0.8694 | 0.75592 | 0.2238 | 37.1639 |
| 8 | BA = [−5.18142 (±1.80169)]<br>+ PEOE_RPC−[46.2538 (±9.44792)]<br>+ a_hyd[0.194747 (±0.0658536)] | 0.9750 | 0.95065 | 0.1051 | 105.963 |
| 9 | BA = [5.54975 (±0.406698)]<br>+ PEOE_VSA_FNEG[2.7359 (±0.731298)]<br>+ SlogP_VSA0[−0.0276773 (±0.00250902)]<br>+ SlogP_VSA4[0.0340621 (±0.0167505)] | 0.9927 | 0.98537 | 0.0600 | 224.618 |
| 10 | BA = [4.54255 (±0.70828)]<br>+ a_don[−0.653515 (±0.0850025)]<br>+ a_hyd[0.187238 (±0.0427888)]<br>+ Inp[−0.222595 (±0.0983743)] | 0.9910 | 0.98215 | 0.0663 | 183.418 |
| 11 | BA = [−5.54731 (±1.26084)]<br>+ PEOE_RPC−[49.2598 (±6.75923)]<br>+ a_hyd[0.166368 (±0.0485236)]<br>+ SlogP_VSA4[0.0353659 (±0.0214624)] | 0.9896 | 0.97935 | 0.0713 | 158.072 |

**Table 3.** Descriptors for quantitative models of HIV-1 RT activity of benzoxazinones

| Compound no. | PEOE_RPC- | PEOE_VSA + 3 | PEOE_VSA_FNEG | a_don | a_hyd | SlogP_VSA0 | SlogP_VSA4 | Inp |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.19346 | 0 | 0.62006 | 1 | 17 | 18.0107 | 8.82159 | 0 |
| **2** | 0.19493 | 0 | 0.60062 | 1 | 16 | 18.0107 | 4.4108 | 0 |
| **3** | 0.19056 | 0 | 0.59449 | 1 | 17 | 18.0107 | 4.4108 | 1 |
| **4** | 0.19127 | 0 | 0.62874 | 1 | 17 | 18.0107 | 8.82159 | 0 |
| **5** | 0.19282 | 0 | 0.49115 | 1 | 15 | 18.0107 | 8.82159 | 0 |
| **6** | 0.19417 | 0 | 0.4781 | 1 | 14 | 18.0107 | 4.4108 | 0 |
| **7** | 0.19014 | 0 | 0.47968 | 1 | 15 | 18.0107 | 4.4108 | 1 |
| **8** | 0.1908 | 0 | 0.51169 | 1 | 15 | 18.0107 | 8.82159 | 0 |
| **9** | 0.17938 | 17.238 | 0.57524 | 2 | 15 | 50.9079 | 8.82159 | 0 |
| **10** | 0.18064 | 17.238 | 0.55826 | 2 | 14 | 50.9079 | 4.4108 | 0 |
| **11** | 0.17688 | 17.238 | 0.55495 | 2 | 15 | 50.9079 | 4.4108 | 1 |
| **12** | 0.1775 | 17.238 | 0.58822 | 2 | 15 | 50.9079 | 8.82159 | 0 |
| **13** | 0.18016 | 8.61901 | 0.50706 | 2 | 16 | 36.0215 | 8.82159 | 0 |
| **14** | 0.17826 | 8.61901 | 0.52562 | 2 | 16 | 36.0215 | 8.82159 | 0 |

**Table 4.** Correlation matrix showing the intercorrelation of molecular descriptors used in models and their correlation with HIV-1 inhibitory activity

| Variables | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* |
|---|---|---|---|---|---|---|---|---|---|
| *A* | 1 | −0.921 | −0.004 | 0.251 | −0.974 | −0.933 | −0.176 | −0.049 | 0.869 |
| *B* | −0.921 | 1 | 0.123 | −0.413 | 0.937 | 0.999 | 0.047 | −0.028 | −0.912 |
| *C* | −0.004 | 0.123 | 1 | 0.618 | 0.01 | 0.112 | 0.115 | −0.083 | 0.241 |
| *D* | 0.251 | −0.413 | 0.618 | 1 | −0.294 | −0.403 | 0.294 | 0.089 | 0.646 |
| *E* | −0.974 | 0.937 | 0.01 | −0.294 | 1 | 0.949 | 0.167 | −0.101 | −0.878 |
| *F* | −0.933 | 0.999 | 0.112 | −0.403 | 0.949 | 1 | 0.06 | −0.036 | −0.914 |
| *G* | −0.176 | 0.047 | 0.115 | 0.294 | 0.167 | 0.06 | 1 | −0.603 | 0.157 |
| *H* | −0.049 | −0.028 | −0.083 | 0.089 | −0.101 | −0.036 | −0.603 | 1 | −0.101 |
| *I* | 0.869 | −0.912 | 0.241 | 0.646 | −0.878 | −0.914 | 0.157 | −0.101 | 1 |

Description of the Table 4:

| | |
|---|---|
| *A* | PEOE_RPC- |
| *B* | PEOE_VSA + 3 |
| *C* | PEOE_VSA_FNEG |
| *D* | a_hyd |
| *E* | a_don |
| *F* | SlogP_VSA0 |
| *G* | SlogP_VSA4 |
| *H* | Inp |
| *I* | $IC_{50}$ (M) |

possessing a partial charge in the range (0.10, 0.15), respectively. An inspection of the correlation matrix (Table 4) showed that these descriptors are highly correlated to each other, which indicates that both the descriptors might represent VSA of the same group of atoms thereby tends to explain the same phenomenon but in a different perspective. Both the models manifest a satisfying statistics and accounted for more than 83% variance in the activity. The *F* statistics of the models are significant at 99% level.

In models 1 and 2 both the P_VSA descriptors has a negative coefficient, which indicates that presence of hydrophilic substituents in the aromatic ring is detrimental to the activity. Introduction of the partial charge descriptor, PEOE_VSA_FNEG in the models 1 and 2 resulted in models 3 and 4 of good statistical quality ($r^2 > 0.97$). It may be observed from Table 4 that the descriptors used for formulating the Eqs. 3 and 4 did not exhibit any colinearity as desired in multivariate analysis. In the models 3 and 4, the coefficient of the descriptor PEOE_VSA_FNEG has a positive weight, which demonstrates the significance of the molecular surface area bearing a fractional negative charge for the HIV-1 RT inhibition by benzoxazinones. The correlation suggests that binding of benzoxazinones to HIV-1 RT involves an electrostatic interaction between electronegative substituents bearing a fractional negative charge in the aromatic ring and a group carrying a fractional positive charge present in the enzyme.

A surprisingly good mono-parametric correlation was obtained with pharmacophore feature descriptor a_don, which refers to the number of hydrogen bond donors present in the molecule (model 5). The model explains 77% of the total variance in the observed activity and the Fischer statistic values indicate 99% level of significance. The coefficient of the descriptor a_don in the model 5 bears a negative sign, which suggests that the presence of hydrogen donor groups such as amino and hydroxyl groups in the aromatic ring is inimical to the activity. The hypothesis is well justified since it was observed earlier in the models 3 and 4 the significance of the groups possessing a fractional negative charge for the activity. The hydrogen donor groups carry a well exposed hydrogen with partial positive charge thereby

will hinder the interaction between the electronegative atom and the electropositive moiety in the enzyme.

Incorporation of another pharmacophore feature descriptor a_hyd into model 5 improved the statistics considerably (model 6). The model shows a positive correlation between the descriptor a_hyd and the HIV-1 RT inhibitory activity of benzoxazinones, which indicates that increase in the hydrophobicity of the molecule, will increase its HIV-1 RT inhibitory potency and hydrophobic interaction is a factor influencing binding of benzoxazinones to HIV-1 reverse transcriptase.

Further statistical analysis by stepwise regression method resulted in mono-parametric and bi-parametric correlations (models 7 and 8). The VSA descriptor PEOE_RPC representing the relative negative partial charge bears a positive coefficient in both the models and positive coefficient of the pharmacophore feature descriptor a_hyd in model 8 supports the conclusions derived from the correlations discussed earlier.

Interestingly, no indicator variable was found in the statistically significant mono-parametric and bi-parametric correlations discussed so far. However, perusal of the structural variations in the benzoxazinone nucleus and the corresponding activity parameters of Table 1 indicate that there might be a significant relationship between the substituents in position $R^1$ and HIV-1 RT inhibition. In the view of above, the study was extended for tri-parametric correlations as they are permitted for a data set of 14 compounds in accordance with the lower limit of rule of thumb.[19] It may be recalled that only regressions allowed are those in which the variables do not intercorrelate with each other.

Model 9 was obtained on addition of another P_VSA descriptor representing lipophilic property SlogP_VSA4 to model 4. The positive coefficient of descriptor SlogP_VSA4 suggests that VSA of atoms contributing to logP (o/w) of the molecule in the range (0.15, 0.20) favours HIV-1 RT inhibition.

Introduction of indicator variable Inp, which refers to *n*-propyl substitution at position $R^1$ into the model 6 resulted in model 10. The negative coefficient of the descriptor in the correlation may possibly highlight steric hindrance caused by the *n*-propyl group. However, the positive coefficient of the pharmacophoric descriptor a_hyd in the model indicates that hydrophobic interactions are involved in the enzyme–drug binding. Thus it may be deduced that the model suggests presence of a hydrophobic pocket that can accommodate bulky substituents at the position $R^1$ but not linear chain substituents. Moreover, the bulky substituents may involve in hydrophobic interaction with the hydrophobic moieties in the enzyme pocket, which explains increase in activity of benzoxazinones with isopropyl substituents at $R^1$.

Model 11 involves PEOE descriptor PEOE_RPC representing the relative partial negative charge of the molecule, pharmacophore feature descriptor a_hyd and P_VSA descriptor describing the lipophilic property SlogP_VSA4. It is worth mentioning that all the tri-parametric correlations accounts for about 98% variance in the biological activity and exhibited better *F* statistics than mono- and bi-parametric correlations.

Finally, in order to estimate the predictive power of the generated correlations, cross validation, following the 'leave-one-out' scheme was performed. The reliability of the correlations was tested in a cross validation with the determination of $r^2$cv (cross validated $r^2$) or $q^2$. The $q^2$ values recorded for the generated correlations are given in Table 5. It is noteworthy that all the correlations exhibit high $q^2$ values and the highest values were recorded for the tri-parametric correlations, which indicate superior predictivity of tri-parametric correlations over the other correlations.

Further confirmation on predictive ability of the correlations was obtained by determining the uncertainty in the prediction ($S_{PRESS}$) and standard error due to prediction (SDEP). The $S_{PRESS}$ and SDEP values should be low for a regression equation to have good predictivity. The $S_{PRESS}$ and SDEP values for the correlations obtained are summarized in the Table 5. The $S_{PRESS}$ and SDEP values were lowest for the tri-parametric correlations and thereby support the aforementioned conclusion. The predicted $\log 1/IC_{50}$ values were calculated using the correlations developed and a comparison was made with the observed values (Table 6).

The aforementioned results and discussion justifies the utilization of the novel set of QuaSAR descriptors for modelling the HIV-1 RT inhibitory activity of benzoxazinones. Another revelation from the study is the utility of the P_VSA descriptors for describing the enzyme–inhibitor interactions as it can be observed from the results of the study that almost 80% of the statistically significant correlations have a P_VSA descriptor. Further, the mono-parametric correlations obtained using P_VSA descriptors were able to explain more than 75% of total variance in the activity and the bi-parametric correlations using a combination of the P_VSA

**Table 5.** Comparison of cross validation parameters for generated QSAR models

| Model no. | $q^{2a}$ | $S_{PRESS}{}^{b}$ | SDEP[c] |
|---|---|---|---|
| 1 | 0.7746 | 0.2151 | 0.1991 |
| 2 | 0.7700 | 0.2173 | 0.2011 |
| 3 | 0.9309 | 0.1244 | 0.1102 |
| 4 | 0.9258 | 0.1289 | 0.1142 |
| 5 | 0.9694 | 0.0867 | 0.0733 |
| 6 | 0.6865 | 0.2537 | 0.2348 |
| 7 | 0.8921 | 0.1554 | 0.1378 |
| 8 | 0.9638 | 0.09440 | 0.0798 |
| 9 | 0.6693 | 0.2605 | 0.2412 |
| 10 | 0.9189 | 0.1347 | 0.1194 |
| 11 | 0.9602 | 0.0990 | 0.0837 |

[a] Squared correlation coefficient of prediction.
[b] Standard deviation of prediction.
[c] Standard error of prediction.

**Table 6.** Experimental ($-\log IC_{50}$) and predicted activity values (models 1–11) for HIV-1 RT inhibitors

| Compound no. | $-\log IC_{50}$ | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 | Model 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 7.11 | 6.76 | 6.76 | 6.95 | 6.94 | 6.77 | 6.82 | 6.92 | 6.96 | 6.82 | 6.88 | 6.86 |
| **2** | 6.89 | 6.77 | 6.77 | 6.94 | 6.94 | 6.77 | 7.08 | 6.67 | 6.99 | 6.83 | 6.89 | 6.84 |
| **3** | 6.81 | 6.75 | 6.74 | 7.03 | 7.02 | 6.75 | 7.01 | 6.68 | 6.97 | 7.11 | 7.10 | 7.02 |
| **4** | 6.99 | 6.79 | 6.79 | 6.59 | 6.6 | 6.79 | 6.62 | 6.82 | 6.65 | 6.70 | 6.70 | 6.79 |
| **5** | 6.68 | 6.81 | 6.80 | 6.57 | 6.58 | 6.81 | 6.41 | 6.93 | 6.50 | 6.49 | 6.48 | 6.46 |
| **6** | 6.56 | 6.81 | 6.81 | 6.59 | 6.61 | 6.82 | 6.67 | 6.67 | 6.54 | 6.51 | 6.44 | 6.46 |
| **7** | 6.52 | 6.79 | 6.78 | 6.66 | 6.67 | 6.79 | 6.63 | 6.69 | 6.54 | 6.77 | 6.69 | 6.64 |
| **8** | 6.7 | 5.86 | 5.85 | 5.91 | 5.92 | 6.03 | 5.99 | 6.08 | 6.03 | 5.98 | 6.03 | 6.09 |
| **9** | 6.09 | 5.96 | 5.95 | 5.95 | 5.96 | 6.10 | 5.86 | 6.21 | 5.96 | 5.86 | 5.91 | 5.89 |
| **10** | 5.72 | 5.94 | 5.93 | 5.91 | 5.92 | 6.08 | 6.04 | 5.98 | 5.95 | 5.8 | 5.82 | 5.81 |
| **11** | 5.82 | 5.87 | 5.86 | 5.97 | 5.98 | 6.04 | 5.99 | 5.96 | 5.92 | 6.05 | 6.04 | 5.99 |
| **12** | 6.05 | 6.31 | 6.35 | 6.19 | 6.16 | 6.00 | 6.18 | 6.11 | 6.28 | 6.24 | 6.24 | 6.32 |
| **13** | 6.22 | 6.30 | 6.34 | 6.25 | 6.21 | 5.98 | 6.15 | 5.96 | 6.17 | 6.28 | 6.20 | 6.17 |
| **14** | 6.32 | 6.73 | 6.73 | 6.95 | 6.95 | 6.73 | 6.97 | 6.79 | 7.07 | 7.02 | 7.05 | 7.13 |

descriptors accounted for more than 95% variance in the activity. The most statistically significant correlations were found to be tri-parametric regressions ($r^2 > 0.95$). Model 9 was found to be the best-fit regression for modelling HIV-1 RT inhibition by benzoxazinones.

Cross validation of the generated models indicated that all the models exhibit good predictivity ($q^2 > 0.6$). The tri-parametric correlations showed almost equivalent predictive ability and high predictive power than the mono-parametric and bi-parametric correlations. The predictive ability of the models was further confirmed from their low $S_{PRESS}$ and SDEP values.

The interpretation of models suggest that presence of substituents having a partial negative charge in the aromatic ring increases the HIV-1 RT inhibitory potency of benzoxazinones and this may be probably due presence of an electropositive group at the site of interaction with the enzyme. Presence of hydrophilic substituents or hydrogen donor groups in the aromatic ring is detrimental to the activity, whereas increase in the number of hydrophobic moieties in the molecule causes a corresponding increase in the inhibitory activity.

## Acknowledgements

Authors wish to thank Tata Elxsi for providing MOE software for the study undertaken. Grateful acknowledgements to Prof. P. B. Sharma, Former Vice Chancellor, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal and Director, SGSITS, Indore for the experimental facilities provided. Authors also thank the honorable referees for their fruitful suggestions, which improved the quality of this paper.

## References and notes

1. De clercq, E. *J. Med. Chem.* **1995**, *38*, 2491.
2. Darke, P. L.; Huff, J. R. In *Advances in Pharmacology*; Augest, J., Ander, M. W., Murad, F., Eds.; Academic: San Diego, 1994; pp 399–415.
3. Miltor, J.; Slater, M. J.; Bird, A. J.; Spinks, D.; Scott, G.; Price, C. E.; Downing, S.; Green, D. V. S.; Madar, S.; Bethell, R.; Stammers, D. K. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2623.
4. METEOR. Press release. Du Pont Pharmaceuticals Company. 18th September, 1998.
5. Young, S. D.; Britcher, S. F.; Tran, L. O.; Rayne, L. S.; Lumma, W. C.; Lyle, T. A.; Huff, J. R.; Anderson, P. S.; Olsen, D. B.; Carroll, S. S.; Pettibone, D. J.; O'Brien, J. A.; Ball, R. G.; Balani, S. K.; Lin, J. H.; Chen, I.-W.; Scheif, W. A.; Sardana, V. U.; Long, W. J.; Byrnes, V. M.; Emine, E. A. *Antimicrob. Agents Chemother.* **1995**, *39*, 2602.
6. Patel, M.; McHugh, R. J.; Cordova, B. C.; Klabe, R. M.; Erickson-Viitanen, S.; Trainor, G. L.; Ko, S. K. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 3221.
7. Lin, A. QuaSAR-descriptors. J. Chem. Comput. Group http://www.chemcomp.com/Journal_of_CCG/Features/descr.htm.
8. MOE is a molecular modeling package developed by Chemical Computing Group Inc., Canada.
9. Labute, P. A. *J. Mol. Graph. Model.* **2000**, *18*, 464.
10. Baurin, N.; Mozziconacci, J.-C.; Arnoult, E.; Chavette, P.; Marot, C.; Morin-Allory, L. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276.
11. Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868.
12. Gasteiger, J.; Marsali, M. *Tetrahedron* **1980**, *36*, 3219.
13. VALSTAT is a statistical program developed by Arun Kumar Gupta, SGSITS, Indore.
14. Kubinyi, H. In *Methods and Principles in Medicinal Chemistry*; Mannhold, R., Krogsgaard-Larsen, R., Timmerman, H., Eds.; VCH: Weinham, 1993; Vol. 1, pp 91–95.
15. *Fundamentals of Mathematical Statistics*; Gupta, S. C., Kapoor, V. K., Eds.; Sulthan Chand and Sons: New Delhi, 1995; pp 3.1–3.25.
16. Kubinyi, H.; Abraham, U. In *3D QSAR in Drug Design, Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers: B.U.: Leiden, 1993; pp 717–728.
17. Muresan, S.; Bologa, C.; Mracec, M.; Chiriac, A.; Jastorff, B.; Simon, Z.; Naráy-Szabó, G. *J. Mol. Struct.* **1995**, *42*, 161.
18. Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Am. Chem. Soc.* **1995**, *117*, 7769.
19. Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem. Lett.* **2001**, *9*, 3287.