

Molecular Surface Features in Modeling the HIV-1 RT Inhibitory Activity of 2-(2,6-Disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones
Ravindra K. Rawal, Yenamandra S. Prabhakar* and S.B. Katti

Medicinal & Process Chemistry Division, Central Drug Research Institute, Lucknow
226001, India

Abstract. The human immunodeficiency virus-1 reverse transcriptase inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones have been analyzed using combinatorial protocol in multiple linear regression (CP-MLR) with several electronic and molecular surface area features of the compounds obtained from Molecular Operating Environment (MOE) software. The study has indicated the role of different charged molecular surface areas in modeling the inhibitory activity of the compounds. The derived models collectively suggested that the compounds should be compact without bulky substitutions on its peripheries for better HIV-1 RT inhibitory activity. It also emphasized the necessity of hydrophobicity and compact structural features for their activity. The scope of the descriptors identified for these analogues have been verified by extending the dataset with different 2-(disubstituted phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones. The joint analysis of extended dataset highlighted the information content of identified descriptors in modeling the HIV-1 RT inhibitory activity of the compounds.

Key Words- QSAR, electronic descriptors, molecular surface area features, HIV-1 RT, non-nucleoside reverse transcriptase inhibitors, 4-Thiazolidinones.

1. Introduction

Reverse transcriptase of the HIV-1 has been successfully exploited as a druggable target for the development of anti-retroviral drugs [1]. The human immunodeficiency virus-1 reverse transcriptase (HIV-1RT) inhibitors mainly fall into two categories viz., nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). These two classes of compounds inhibit through mutually exclusive mechanisms. The NRTIs are irreversible inhibitors and inhibit the enzyme activity by covalently binding to the active site. Where as NNRTIs elicit their activity by reversibly binding to the enzyme's allosteric site through non-covalent interaction. Between these two classes of inhibitors, NNRTIs have proved to be more specific with little cytotoxicity [2]. Because of this significant research efforts have been focused on different chemical scaffolds as NNRTIs [3]. The availability of vast chemical diversity in HIV-RT inhibitors coupled with the X-ray crystal data of enzyme-inhibitor complexes opened avenues for QSAR and modeling studies of these inhibitors to obtain better insight into their chemical features [1,2,3]. Moreover, the X-ray crystallographic studies of diverse NNRTIs/RT complexes have indicated that irrespective of the structure class involved, the compound's ability to assume "butterfly-

* Corresponding author. Tel.: +91-522-2612411; fax: +91-522-2623405
E-mail address: yenpra@yahoo.com (Y.S. Prabhakar)

like” shape is vital for the binding and inhibition of HIV-1 RT [4]. The QSAR and modeling studies on the HIV-1 RT inhibitory activity of 2, 3-diaryl-thiazolidin-4-ones by our group as well as others indicated the importance of overall hydrophobicity of the analogues, and steric and electronic features of meta- / para- substituents of 3-aryl moiety [5,6,7]. It also suggested that a heteroaryl system would be a preferred one for the 3-aryl moiety for better HIV-1 RT inhibitory activity [8,9]. The studies carried out with the topological descriptors of these compounds suggested that less extended or compact saturated structural templates would be better for the activity. In this, the structural fragments of lengths seven and four were found to be associated with activity information [9]. To widen the understanding of feature space of the HIV-1 RT inhibitors, in this investigation we attempted the structure-activity relations with their electronic properties.

2 Materials and Methods

2.1 Dataset

2-(2,6-Disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (referred to as pyrimidine analogues) (Table 1) [10-12] and 2-(disubstituted phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (referred to as pyridine analogues) (Table 2) [8,11-14] along with their reported HIV-1 RT inhibitory activity (EC_{50} , effective concentration, in moles per liter, to reduce HIV-1 induced cytopathic effect by 50% in MT-4 cells) expressed in terms of logarithm of inverse of inhibitory concentration ($-\log EC_{50}$) have been considered for the analysis. In this study, for the generation of 3D-structures of the analogues, the structures of compounds 1-3 of Table 1 and compounds 46-48 of Table 2 have been considered as the common basic core templates. These core structures (compounds 1-3 and 46-48) have been generated in Molecular Operating Environment (MOE) [15,16] using the standard template library of the software followed by systematic conformational search under default conditions with energy minimization (force field: MMFF94x) to identify the minimum energy conformer. The structures of all other analogues of Tables 1 and 2 have been generated by appropriately modifying the corresponding core structures and subjected them to energy minimization (force field: MMFF94x) to obtain their final conformation. For the QSAR study, the molecular descriptors were calculated in MOE using these conformations. As the computation of 3D-descriptors take the cartesian coordinates into consideration, these descriptor values are sensitive to changes in the atomic positions. Table 3 presents the scope of all descriptor classes in describing the molecular structure. Our earlier experience with the HIV-1 RT inhibitors suggest that non-polar substituent of 3-aryl moiety of 2,3-diaryl-1,3-thiazolidin-4-ones have influence on the activity [5]. Accordingly, an indicator parameter (I_1) has been added to the dataset to address the nature of 3-aryl substituents (for non-polar groups/atoms (Me/H), $I_1=1$ and ‘zero’ otherwise). As the total number of descriptors involved in this study is very large, 187 parameters, only the names of descriptor classes and the participating descriptors have been addressed in the discussion. The QSAR model generation and validation have been done using the combinatorial protocol in multiple linear regression (CP-MLR) [17]. The computational procedure is briefly described below.

2.2 Computational Procedure

CP-MLR is a ‘filter’ based variable selection procedure for model development in QSAR and QSPR studies [5,17]. The detailed procedure is discussed in the earlier reports. Briefly, in this procedure a combinatorial strategy with appropriately placed ‘filters’ has

been interfaced with MLR to result in the extraction of diverse structure-activity models, in the predefined confines of descriptor, each having unique combination of descriptors from the dataset under study. In this, the filters address inter-parameter correlation cutoff criteria for variables in a subset regression (filter-1, default value 0.3), t-values of the regression coefficients of the variables (filter-2, default value 2.0), internal explanatory power of the models with different number of variables expressed as square-root of adjusted multiple correlation coefficient of regression equation, \bar{r} (filter-3, default value 0.74) and the external consistency of the variables of the model in the form of cross-validated R^2 (Q^2) criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default limits $0.3 \leq Q^2 \leq 1.0$). In CP-MLR, filter-1 proposes tentative variable subsets (seeds) and other filters evaluate their significance (fitness criteria) in explaining the activity under investigation. The composition of variable subsets to be evaluated for their fitness can be altered by assigning different threshold values to filter-1. In this, a default value of 0.3 has been assigned to filter-1 as it gives almost orthogonal variable seeds for fitness evaluation. Nevertheless, filter-1 can be assigned with higher threshold values to enable to evaluate larger number of variable seeds, where necessary. Of course, the increased threshold level of filter-1 greatly increases the computational time for finding optimum solutions. It may be stressed that filter-1 only proposes the variable seeds whereas the criteria of their fitness in models are solely decided by the other filters. In addition to cross-validation, each identified model has been reassessed for the chance correlations, if any, by repeated randomization of the biological response [5,18]. The emerging regression equations with correlation coefficients better than or equal to the one corresponding to unscrambled response data have been counted to express the percent chance correlation of the model under examination. For this every model has been subjected to one hundred such simulation runs with scrambled activity. The model development procedure has been further verified by creating two divergent test sets with each containing about one third of the total compounds under analysis. For this, one test set has been formed from the bit-packed version of the MACCS fingerprints (FP BIT MACCS) [15] of the compounds and the other from the random selection procedure. Besides CP-MLR, stepwise regression procedure of the SYSTAT [19] has been used for the analysis of these analogues.

3. Results and Discussion

The HIV-1 RT inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (Table 1) [10-12] have been analyzed in CP-MLR (filter-1 ≤ 0.3 , filter-2 ≥ 2.0 , filter-3 ≥ 0.74 and filter-4 as $0.3 \leq Q^2 \leq 1.0$) for one to three descriptor models in terms of physicochemical/ topological (2D) and atomic/ group electronic (3D) properties (Table 3) of the compounds. This has led to the identification of a three-parameter equation for the activity with petitjean descriptor (petitjean), total polar surface area (TPSA) and the y component of the dipole moment (dipoleY) of the compounds as shown.

$$-\log EC_{50} = -7.879 + 30.015(7.755)\text{petitjean} - 0.078(0.022)\text{TPSA} - 4.194(1.033)\text{dipoleY}$$

$$n = 21, r = 0.787, Q^2 = 0.432, s = 0.644, F = 9.217 \quad (1)$$

In this and in all other regression equations, n is the number of compounds, r is the correlation coefficient, Q^2 is cross-validated R^2 from leave-one-out (LOO) procedure, s is the standard error of the estimate and F is the F-ratio between the variances of calculated

and observed activities. The values given in the parentheses are the standard errors of the regression coefficients. As there are no more equations up to 3 descriptors within the interparameter correlation limits of 0.3 ($\text{filter-1} \leq 0.3$), the study has been extended by relaxing the filter-1 to 0.74. It may be mentioned that, this increase in the threshold value of filter-1 alters the composition of variable subsets (seeds) to be evaluated for their fitness to explain the activity. The value 0.74 has been chosen to balance number of variable seeds and computational time for evaluation. Highly correlated variables in regression models may cause uncertainty in the coefficients of the corresponding variables. However, keeping this in view care has been taken to prevent the accumulation of spurious equations through validation and randomization strategies. In CP-MLR, the relaxation of filter-1 to 0.74 has resulted in the identification of 26 regression equations. Of these, 2 are two-descriptor models and 24 are three-descriptor models. The essence of the descriptors identified in all these equations has been provided in Table 4 in the form of their average of regression coefficients, along with standard deviation across the models and the total incidence corresponding to all the equations. This, while providing the averages of the estimated regression coefficients of all the identified descriptors, shows their variance across the models emerged from the study. Equations 2 and 3 are some selected three-parameter models for the HIV-1 RT inhibitory activity of the compounds. The four-parameter equations (equations 4 and 5) shown here are part of a large number of higher descriptor models that exist in the 27 selected descriptors of Table 4.

$$-\log EC_{50} = -25.539 + 64.302(9.839)\text{petitjean} + 0.538(0.151)a_nCl + 2.951(0.482)I_1$$

$$n = 21, r = 0.853, Q^2 = 0.609, s = 0.545, F = 15.15 \quad (2)$$

$$-\log EC_{50} = -35.793 + 57.424(8.511)\text{petitjean} + 15.112(2.961)Q_VSA_FHYD + 4.943(0.847)\text{dipole}$$

$$n = 21, r = 0.868, Q^2 = 0.638, s = 0.519, F = 17.24 \quad (3)$$

$$-\log EC_{50} = -36.496 + 58.485(6.926)\text{petitjean} + 15.518(2.410)Q_VSA_FHYD + 5.237(0.695)\text{dipole} - 0.027(0.009)(PEOE_VSA+2)$$

$$n=21, r=0.920, Q^2 = 0.639, s = 0.422, F = 22.004 \quad (4)$$

$$-\log EC_{50} = 3.912 + 0.0744(0.010)vsa_hyd - 0.022(0.008)(ASA+) - 0.009(0.001)DCASA - 3.392(0.731)std_dim1$$

$$n = 21, r = 0.915, Q^2 = 0.657, s = 0.434, F = 20.528 \quad (5)$$

In the randomization study (hundred simulations per model) none of the identified models have shown any chance correlation. The plots of predicted and fitted versus observed activity of all equations are shown in Figure 1. The equations are further validated through two test sets, each containing seven compounds out of the twenty-one active analogues listed in Table 1. Of these two test sets, one has been constructed from the cluster analysis of the MACCS fingerprints (FP BIT MACCS) of the twenty-one active compounds (Table 1). As all the compounds are structurally closely related, 85% similarity level has been used to cluster them into different groups. The second test set of the compounds corresponds to the random selection procedure. The predictions of the test sets have been done with the models, corresponding to equations 2-5, developed using the fourteen compounds remained in the training groups. The statistics associated with

the training and test set are shown in Table 5. The predictions made for the both test sets are within the reasonable limits of their experimental values (Table 1). In the literature compounds 8,9,17 and 18 (Table 1) are simply mentioned as 'not active' in the test system [10-12]. Interestingly, majority of the models estimated the activity profile of these 'not active' compounds as lower than the least active compound included in the regression, which means that they are ranked lower than the lowest active compound of the series. This overall satisfactory performance of the training set equations in predicting the activity of test sets lent confidence to the variables and the models selected in the study.

A summary of the physical meaning of the participating descriptors in relation to the activity is presented in Table 6. A brief discussion of these descriptors is presented in the following paragraphs. The regression coefficient of petitjean descriptor (petitjean in equations 1-4) suggests that compounds with compactly connected atoms would lead to better activity. Similar to petitjean descriptor, in some other models the descriptors rgyr (radius of gyration) and pmiY (y component of the principal moment of inertia) (Table 4) have taken part with the negative coefficients. All these parameter collectively convey that bulky substitution away from the center of the molecule is not favorable for the activity. Moreover, in equation 5, the descriptor std_dim1 (standard dimension 1) corresponds to the largest eigenvalue of the covariance matrix of the atomic coordinates and represents the standard deviation along a principal component axis. The regression coefficient of this descriptor suggests in favor of smaller eigenvalue for better activity. As these eigenvalues are from the covariance matrix of the atomic coordinates, compact or closely arranged atoms of the compounds would lead to desired values. This is in concurrence with the regression coefficients of petitjean and rgyr descriptors.

Several properties of the van der Waals surface area (VSA) of these compounds have taken part in the regression equations. Among these, the descriptors prefixed with PEOE have resulted from the partial equalization of orbital electronegativities (PEOE) method of calculating atomic partial charges and the descriptors prefixed with Q have resulted from the partial charges on the atoms of the structure. Q_VSA_FHYD in equations 3 and 4 represents the fraction of the hydrophobic surface area with partial charge less than or equal to 0.2. Its positive regression coefficient suggests that increase in the compound's hydrophobic surface area would be favorable for the activity. The regression coefficient of vsa_hyd (sum of VSA of hydrophobic atoms) in equation 5 also suggests the same. In equation 4, the descriptor (PEOE_VSA+2) represents the sum of VSA with partial atomic charges in the range of 0.10 to 0.15. It has participated in the model with a negative regression coefficient. This and other related descriptors suggest in favor of decreased positively charged and increased negatively charged surface area for better inhibitory activity.

In equation 5, the descriptors (ASA+) and DCASA account for the partially positively charged atoms' total water accessible surface area (ASA) and the absolute difference between the positive and negative partial charge-weighted ASA, respectively, of the compound. The negative regression coefficients of these descriptors suggest the unfavorable conditions associated with these kinds of positively charged and polar atomic regions for the activity. The participation of dipole moment (dipole) in equation 3 suggests in favor of some molecular polarity for the activity. The participation of the y

component of the dipole moment (dipoleY) in equation 1 suggests the nature and directionality of this component in the compound's interaction with the receptor.

In equation 2, the parameters a_{nCl} and I_1 represent the count of chloro groups and the apolar 3-pyrimidinyl substitution (H or methyl), respectively, of these compounds. The positive regression coefficient of I_1 favors apolar environment in the vicinity of 3-pyrimidinyl moiety for the better inhibitory activity. The a_{nCl} feature (equation 2) has been further investigated with a site specific indicator parameter (I_2) in terms of number of chloro group at 2'- and 6'-positions of 2-aryl moiety of the thiazolidin-4-one (equation 6).

$$-\log EC_{50} = -23.469 + 60.092(8.853)petitjean + 0.573(0.143)I_2 + 2.760(0.426)I_1$$

$$n = 21, r = 0.869, Q^2 = 0.684, s = 0.516, F = 17.573 \quad (6)$$

A comparison of equations 2 and 6 indicate that the model incorporating I_2 (equation 6) in place of a_{nCl} (equation 2) has resulted in overall improved statistics. Besides these equations, several other models exist among the descriptors presented in the Table 4 with one or more parameters as common descriptors to equations 1-5. In brief, all the identified descriptors collectively suggest that for better HIV-1 RT inhibitory activity the compounds should be compact without bulky substitutions on its peripheries. The dipoleY suggested the directionality component in the interaction with the receptor. The activity prefers and tolerates the hydrophobic and negatively charged surfaces of the compound.

The 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (Table 1) [10-12] have also been jointly analyzed with 2-(disubstituted phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Table 2) [8,11-14] for their HIV-1 RT inhibitory activity. It has been undertaken with a view to examine the scope of the 27 descriptors (Table 4) identified for 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (Table 1) in modeling the activity of a larger dataset. The following equation is a stepwise regression model of the combined set of compounds (Tables 1 and 2) from the 27 descriptors (Table 4).

$$-\log EC_{50} = -14.939 + 0.092(0.019)(PEOE_VSA+2) - 0.040(0.010)(PEOE_VSA-1)$$

$$+ 24.698(6.227)PEOE_VSA_FNEG - 0.052(0.012)Q_VSA_NEG$$

$$+ 0.086(0.017)Q_VSA_POL - 0.189(0.027)TPSA - 1.100(0.327)\text{LogP}(O/W)$$

$$+ 5.213(0.927)\text{dipole} + 0.037(0.015)(ASA+) + 3.848(0.861)\text{std_dim2}$$

$$+ 2.676(0.561)\text{dipoleY} - 2.241(0.389)I_1 + 2.871(0.408)I_2$$

$$n = 75, r = 0.811, Q^2 = 0.423, s = 0.544, F = 9.003 \quad (7)$$

This as well carries two test sets, with 25 compounds each, created as described in the case of compounds of Table 1. The test sets predictions by the corresponding training set equations are satisfactory (Tables 1 and 2). When all the descriptors (187 descriptors) are considered for the analysis, another 13-parameter equation has been found for these compounds (Tables 1 and 2) with a correlation coefficient of 0.914 (equation not shown). Of these 13 descriptors, only one (std_dim2) is common with the 27 descriptors listed in Table 4. However, as outlined earlier, the purpose of this exercise is to examine the scope of the 27 descriptors identified for the compounds of Table 1 in explaining the activity beyond this dataset. From the stepwise regression of 27 descriptors, equation 7 is the simplest one that we could find for the compounds of Tables 1 and 2 together. The significance of this equation reduces with lesser number of descriptors. Addition of more

variables to this equation did not improve the results. Attempts with partial least squares analysis also did not yield any better model. The descriptors of this equation draw the same inference as discussed in the case of pyrimidine analogues (Table 1) alone. It suggests that these descriptors (Table 4) hold at least some modeling information beyond the pyrimidine analogues for which they were identified. This provides some scope to make use of these descriptors to rank the activity of other related compounds.

4. Conclusions

The study has provided detailed structure activity relationship of the HIV-1 RT inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (Table 1) and 2-(disubstituted phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Table 2) in terms of different charge, polar and hydrophobic surface area requirements. The activity of the compounds is a cumulative influence of different structural features which are identified in terms of individual descriptors. The positive regression coefficient of parameters a_{nCl} and I_2 clearly suggest in favor of chloro groups on the 2-aryl moiety of these molecules for the better activity. Also, the regression coefficients of descriptors ‘dipole’ and ‘dipoleY’ in these compounds suggest in favor of large dipole moment for better activity. Additionally, the descriptor ‘dipoleY’ has set a directional requirement for the activity. These features are clearly visible in compounds 4, 5, 10, 11 (high active compounds where 2-aryl is substituted with halo groups and 3-aryl is substituted with methyl only) versus compound 7 (low active compound where 2-aryl is substituted with halo groups and 3-aryl is substituted with chloro and methyl groups) with reduced the dipole moment. In addition this, the regression coefficient of indicator I_1 , defined based on our earlier experience with 2,3-diaryl thiazolidin-4-ones [5], also advocate non-polar substituent groups for the 3-aryl moiety. In terms of partially charged surface areas, the thiazolidin-4-one scaffold has provided partially negatively charged surface area and its 2- and 3-aryl moieties have provided the partially positively charged surface areas in these compounds. In high active compounds (e.g. compound 10), the partially positively charged surface area is more prominent on 2-aryl moiety. This phenomenon versus activity is in agreement with the coefficients of other descriptors such as ‘dipole’, ‘dipoleY’, a_{nCl} , I_1 and I_2 . The activity prefers and tolerates the hydrophobic surfaces of the compound. All the identified descriptors collectively suggested that the compounds should be compact without bulky substitutions on its peripheries for better HIV-1 RT inhibitory activity. Our earlier studies on different 2,3-diaryl thiazolidin-4-ones as HIV-1 RT inhibitors also indicated a preference for hydrophobic compounds with less extended or compact structural templates for better activity [5,6]. The indicator parameter I_1 of this study is a result of past experience with 2,3-diaryl thiazolidin-4-ones [5]. The analyses carried out on the training/ test sets have given additional support to the findings of this study. The joint analysis of pyrimidines (Table 1) and pyridines (Table 2) suggest that the descriptors identified in Table 4 hold at least some modeling information beyond the pyrimidine analogues and provide some scope to rank the activity of other related compounds. These results will be helpful to advance the understanding of modeling aspects of 4-thiazolidinones as HIV-1 RT inhibitors.

Supporting information: Cartesian coordinates of compound 10 in PDB format and all descriptors and SMILES notation of compounds.

Acknowledgements

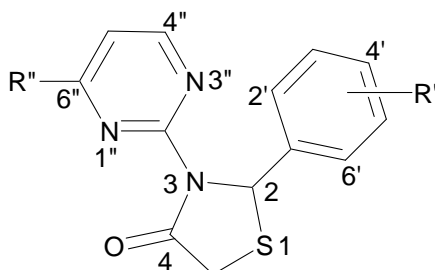
The authors thank the Director, CDRI for the interest in the work. CDRI Communication No.6852

References

1. H. Jonckheere, J. Anne, E. De Clercq, *Med. Res. Rev.* **2000**, *20*, 129-154.
2. E. De Clercq, *J. Med. Chem.* **2005**, *10*, 1297-1313.
3. R. Garg, S. P. Gupta, H. Gao, M. S. Babu, A. K. Debnath, C. Hansch, *Chem. Rev.* **1999**, *99*, 3525-3601.
4. W. Schafer, W. G. Friebe, M. Leinert, A. Merttens, T. Poll, W. Von der Saal, H. Zilch, H. Nuber, M. L. Ziegler, *J. Med. Chem.* **1993**, *36*, 726-732.
5. Y. S. Prabhakar, V. Raja. Solomon, R. K. Rawal, M. K. Gupta, S. B. Katti, *QSAR Comb. Sci.* **2004**, *23*, 234-244.
6. (a) R. K. Rawal, V. Raja. Solomon, Y. S. Prabhakar, S. B. Katti, E. De Clercq, *Combinatorial Chemistry and High-throughput screening.* **2005**, *8*, 439-443. (b) R. K. Rawal, Y. S. Prabhakar, S. B. Katti, E. De Clercq, *Bioorg. Med. Chem.* **2005**, *8*, 439-443.
7. K. Roy, T. Leonard, *QSAR Comb. Sci.* **2005**, *24*, 579-592.
8. M. L. Barreca, J. Balzarini, A. Chimirri, E. De Clercq, L. D. Luca, M. H. Holtje, M. Holtje, A. M. Monforte, P. Monforte, C. Pannecouque, A. Rao, M. Zappala, *J. Med. Chem.* **2002**, *45*, 5410-5413.
9. Y. S. Prabhakar, R. K. Rawal, M. K. Gupta, V. Raja Solomon, S. B. Katti, *Combinatorial Chemistry and High-throughput Screening.* **2005**, *8*, 431-437.
10. A. Rao, J. Balzarini, A. Carbone, A. Chimirri, E. De Clercq, A. M. Monforte, P. Monforte, C. Pannecouque, M. Zappala, *Antiviral Res.* **2004**, *63*, 79-84.
11. A. Rao, J. Balzarini, A. Carbone, A. Chimirri, E. De Clercq, A. M. Monforte, P. Monforte, C. Pannecouque, M. Zappala, *Il Farmaco* **2004**, *59*, 33-39.
12. A. Rao, A. Carbone, A. Chimirri, E. De Clercq, A. M. Monforte, P. Monforte, C. Pannecouque, M. Zappala, *Il Farmaco* **2002**, *57*, 747-751.
13. M. L. Barreca, J. Balzarini, A. Chimirri, L. D. Luca, A. M. Monforte, P. Monforte, A. Rao, M. Zappala, J. Balzarini, E. De Clercq, C. Pannecouque, M. Witvrouv, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1793-1796.
14. M. L. Barreca, A. Chimirri, E. De Clercq, L. D. Luca, A. M. Monforte, P. Monforte, A. Rao, M. Zappala, *Il Farmaco* **2003**, *58*, 259-263.
15. MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University Street, Suite 1600, Montreal, Quebec, Canada H3B 3X3
16. www.chemcomp.com; www.chem.ac.ru/chemistry/soft/moperenv.en.html.
17. Y. S. Prabhakar, *QSAR Comb. Sci.* **2003**, *22*, 583-595.
18. S. S. So, M. Karplus, *J. Med. Chem.* **1997**, *40*, 4347-4359.
19. SYSTAT, Version 7.0: SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611.

Legend for figure 1

Figure 1. Plots of the predicted (O) and fitted (X) versus observed HIV-1 RT inhibitory activities of the 4-thiazolidinones for equations 2-7.

Table 1. HIV-1 RT inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones.


Comp No	R''	R'		-logEC ₅₀															
		2'	6'	Obs ^a	Eq.2 ^b	Set1 ^c	Set2 ^c	Eq.3	Set 1	Set 2	Eq.4	Set 1	Set 2	Eq.5	Set 1	Set 2	Eq.7	Set 1	Set 2
1	H	Cl	Cl	6.64	7.16	7.07	7.01	6.45	6.44	6.34	6.77	6.80	6.71	6.43	6.45	6.49	6.71	6.56	6.65
2	H	Cl	F	6.24	6.55	6.62	6.55	6.10	6.07	6.01	6.13	6.15	6.08	6.30	6.18	6.21	6.39	6.24	6.36
3	H	F	F	5.44	6.06	6.08	6.18	5.97	5.92	5.84	5.75	5.73	5.66	4.67	4.68	4.71	6.20	6.15	6.33
4 ^d	H	Cl	Cl	7.36	7.01	7.07	6.79	6.98	6.98	6.82	7.34	7.38	7.26	7.56	7.68	8.06	7.46	7.41	7.53
5	H	Cl	F	7.13	6.49	6.51	6.55	6.63	6.58	6.57	6.70	6.71	6.68	7.29	7.40	7.33	7.16	7.12	7.26
6	H	F	F	6.41	5.94	6.08	6.13	6.56	6.52	6.44	6.36	6.38	6.30	6.36	6.29	6.22	7.07	7.10	7.13
7	Cl	Cl	Cl	4.29	5.05	5.04	4.89	4.76	4.99	4.95	4.63	5.00	4.93	4.30	4.14	4.39	6.29	5.86	6.51
8 ^e	Cl	Cl	F	-	4.12	4.14	4.20	4.35	4.40	4.37	4.03	4.14	4.12	4.04	3.89	3.92	5.55	5.06	5.31
9 ^e	Cl	F	F	-	3.58	3.64	3.82	4.14	4.20	4.16	3.57	3.66	3.66	3.74	3.49	3.46	5.37	4.89	5.22
10	Me	Cl	Cl	7.77	6.92	6.77	6.92	7.10	6.99	7.11	7.46	7.49	7.50	6.98	7.10	7.20	6.65	6.53	6.61
11	Me	Cl	F	7.42	6.47	6.48	6.45	6.84	6.76	6.68	6.90	6.92	6.83	6.81	6.84	6.69	6.50	5.98	6.56
12	Me	F	F	6.77	5.89	5.86	6.06	6.80	6.75	6.67	6.59	6.61	6.55	6.67	6.60	6.38	6.55	6.17	6.77
13	MeO	Cl	Cl	7.62	7.74	7.76	7.25	7.34	7.60	7.07	7.73	7.89	7.58	6.98	7.14	6.97	6.86	6.24	6.70
14	MeO	Cl	F	7.25	7.10	7.11	7.10	7.00	7.27	7.00	6.99	6.98	6.99	7.00	7.07	6.98	6.74	6.81	6.70
15	MeO	F	F	6.58	6.64	6.69	6.95	7.12	7.11	7.13	6.75	6.70	6.62	6.97	6.89	6.74	6.71	6.64	6.74
16	OH	Cl	Cl	4.48	3.72	3.72	3.88	4.19	4.20	4.04	4.50	4.54	4.36	4.56	4.28	4.51	6.55	6.52	6.98
17 ^e	OH	Cl	F	-	3.58	3.64	3.82	3.91	3.94	3.87	3.84	3.84	3.77	4.47	4.22	4.27	5.41	5.07	5.54

18 ^e	OH	F	F	-	3.04	3.14	3.45	3.76	3.79	3.71	3.45	3.42	3.36	4.45	4.12	4.10	5.27	4.98	5.48
19	H	Me	Me	5.47	6.06	6.08	6.18	5.58	5.57	5.53	5.87	5.89	5.82	6.20	6.07	6.28	5.64	5.54	5.66
20	H	MeO	F	6.27	5.96	6.08	6.16	5.63	5.69	5.43	5.68	5.71	5.45	6.25	6.23	6.32	5.47	5.37	5.48
21 ^f	H	F	CF ₃	6.04	6.56	6.64	6.55	5.15	5.03	5.25	5.20	5.00	5.22	6.75	7.34	6.98	6.59	6.50	7.44
22 ^g	H	Cl	F	6.11	6.56	6.63	6.60	6.97	6.98	6.91	7.04	7.17	7.09	7.55	7.71	7.21	6.41	6.13	6.33
23	H	MeO	MeO	6.04	5.98	6.09	6.20	6.83	6.84	6.62	7.75	7.76	7.61	5.35	4.74	5.59	6.04	5.99	6.09
24 ^h	H	Cl	Cl	6.55	7.17	7.30	7.04	6.66	6.62	6.53	6.19	6.17	6.12	6.71	6.87	7.11	5.50	5.13	5.40
25 ^h	Me	F	F	5.19	6.09	6.35	6.37	6.54	6.54	6.44	5.96	6.24	6.02	5.80	5.69	5.54	4.88	4.66	5.04

^a, Observed activity; ref 10, compounds 1-18; ref 11, compounds 19-23; ref 12, compounds 24-25;

^b, for all equations 'set 1' corresponds to MACCS cluster and 'set 2' corresponds to random selection, unless otherwise stated all predicted activities are from LOO cross validation;

^c, test set compounds are identified in italics and their activities are predicted with the corresponding model developed using the remaining compounds as the training group;

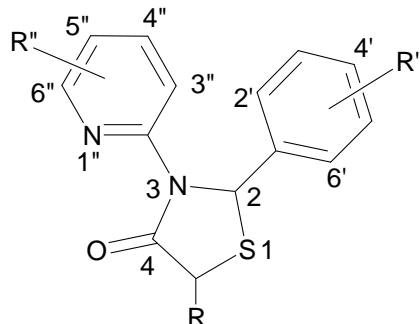
^d, compounds 4-25 are with methyl group at 4'' position;

^e, not active;

^f, in compound 21, 3' position is substituted with chloro group;

^g, in compound 22, 3' position is substituted with methyl group;

^h, thiazolidin-4-(thi)-one in place of thiazolidin-4-one.

Table 2. HIV-1 RT inhibitory activity of 2-(disubstituted phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones.


Comp No	R''			R'			R	-logEC ₅₀			
	4''	5''	6''	2'	3'	6'		Obs ^a	Eq.7 ^b	Set1 ^c	Set2 ^c
26	Me	H	H	Me	H	Me	H	5.24	4.45	4.55	4.10
27	Me	H	H	F	H	MeO	H	6.06	6.67	6.52	7.24
28	Me	H	H	F	Cl	CF ₃	H	5.67	5.35	5.13	5.75
29	Me	H	H	Cl	Me	F	H	6.49	6.45	6.48	6.36
30	Me	H	H	MeO	H	MeO	H	5.48	5.73	5.76	5.73
31	H	H	Me	Me	H	Me	H	5.81	4.99	4.77	4.76
32	H	H	Me	F	H	MeO	H	6.54	6.75	6.96	6.90
33	H	H	Me	F	Cl	CF ₃	H	5.52	5.63	5.52	5.91
34	H	H	Me	Cl	Me	F	H	7.30	6.83	6.93	6.82
35	H	H	Me	MeO	H	MeO	H	5.87	6.09	6.23	6.14
36	H	H	Br	Me	H	Me	H	5.20	5.48	5.44	5.59
37	H	H	Br	F	H	MeO	H	7.47	7.65	8.08	7.92
38	H	H	Br	F	Cl	CF ₃	H	6.18	6.20	6.12	6.63
39	H	H	Br	Cl	Me	F	H	6.46	6.48	6.48	6.50
40	H	H	Br	MeO	H	MeO	H	7.35	5.72	5.60	5.61
41	Me	H	Me	Me	H	Me	H	5.36	5.14	5.12	5.10
42	Me	H	Me	F	H	MeO	H	6.70	6.30	6.51	6.20
43	Me	H	Me	F	Cl	CF ₃	H	5.23	6.02	5.99	6.18
44	Me	H	Me	Cl	Me	F	H	6.63	6.90	7.02	6.82
45	Me	H	Me	MeO	H	MeO	H	5.95	6.66	7.20	6.83
46	H	H	H	Cl	H	Cl	H	6.75	6.57	6.56	6.52
47	H	H	H	Cl	H	F	H	6.56	6.37	6.42	6.31
48	H	H	H	F	H	F	H	6.07	6.11	6.16	6.14
49	H	Cl	H	Cl	H	Cl	H	5.75	5.16	4.94	4.89
50	H	Cl	H	Cl	H	F	H	5.67	5.36	5.35	4.78
51	H	Cl	H	F	H	F	H	5.14	5.14	5.08	4.87
52	H	Br	H	Cl	H	Cl	H	5.82	6.02	5.90	5.92
53	H	Br	H	Cl	H	F	H	5.91	5.92	5.87	5.90
54	H	Br	H	F	H	F	H	5.31	6.04	6.09	6.19
55	H	H	Br	Cl	H	Cl	H	6.57	7.16	7.33	7.18
56	H	H	Br	Cl	H	F	H	7.19	6.75	6.83	6.92
57	H	H	Br	F	H	F	H	7.52	6.78	6.73	7.06

58 ^{d,e}	H	H	H	Cl	H	Cl	H	NA	5.60	5.45	5.35
59 ^e	H	H	H	Cl	H	F	H	4.27	5.30	5.24	5.28
60	Me	H	H	Cl	H	Cl	H	6.83	6.68	6.64	6.65
61	Me	H	H	Cl	H	F	H	7.00	6.49	6.48	6.45
62	Me	H	H	F	H	F	H	6.61	6.16	6.12	6.21
63	H	Me	H	Cl	H	Cl	H	5.85	6.98	7.18	6.92
64	H	Me	H	Cl	H	F	H	5.29	6.37	6.40	6.37
65	H	H	Me	Cl	H	Cl	H	7.36	7.24	7.39	7.27
66	H	H	Me	Cl	H	F	H	7.28	7.12	7.28	7.21
67	H	H	Me	F	H	F	H	7.09	6.73	6.81	6.86
68	Me	H	Me	Cl	H	Cl	H	7.05	7.22	7.25	7.27
69	Me	H	Me	Cl	H	F	H	7.38	6.99	7.05	7.04
70	Me	H	Me	F	H	F	H	7.23	6.87	6.95	7.03
71 ^f	H	H	Me	F	H	F	H	5.19	5.27	5.24	5.24
72 ^f	H	H	Me	Cl	H	Cl	H	4.94	5.66	5.75	5.52
73 ^f	Me	H	H	F	H	F	H	5.36	5.17	5.08	5.05
74 ^{d,f}	Me	H	H	Cl	H	Cl	H	NA	5.53	5.51	5.35
75 ^f	H	H	Br	F	H	F	H	6.24	5.97	6.13	6.11
76 ^g	H	H	Me	F	H	F	Me	5.47	5.66	5.88	5.60
77 ^g	H	H	Me	Cl	H	Cl	Me	5.42	5.40	5.37	5.02
78 ^g	Me	H	Me	F	H	F	Me	5.51	6.74	6.80	6.99
79 ^h	H	H	Me	F	H	F	Me	5.72	5.27	5.38	5.16
80 ^h	H	H	Me	Cl	H	Cl	Me	5.88	5.27	5.17	5.12
81 ^h	Me	H	Me	F	H	F	Me	5.69	5.88	6.22	5.82

^a, Observed activity; ref 11, compounds 26-45; ref 8, compounds 46-70; ref 12, compounds 71-81;

^b, see foot note 'b', Table 1;

^c, see foot note 'c', Table 1;

^d, not active;

^e, in compounds 58-59, 3'' position is substituted with methyl group;

^f, thiazolidin-4-(thi)one in place of thiazolidin-4-one;

^g, *trans*-isomer;

^h, *cis*-isomer.

Table 3. Definitions and scope of descriptors classes used in the study [15].

Descriptor Class	Definition and scope
2D	Conformation independent molecular descriptors from connection table of a molecule (<i>e.g.</i> , elements, formal charges and bonds).
i3D (Internal 3D)	Descriptors from 3D coordinate information of molecule but invariant to rotations and translations of the conformation.
x3D (External 3D)	Descriptors from 3D coordinate information with an absolute frame of reference

Table 4. Descriptors identified to model the HIV-1 RT inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones (**Table 1**) along with their average regression coefficients.

Descriptor ^a	Av Reg Coef (sd) Total incidence ^b	Descriptor ^a	Av Reg Coef (sd) Total incidence ^b
2D		I_1	2.582(0.273)8
petitjean	54.505(9.794)16	i3D	
a_nCl	0.538(0.000)1	dipole	4.941(0.867)6
PEOE_VSA+2	-0.029(0.000)1	rgyr	-3.762(0.000)1
PEOE_VSA-1	0.017(0.000)1	ASA+	-0.018(0.000)1
PEOE_VSA_FNEG	8.515(2.916)2	ASA_P	-0.024(0.000)1
PEOE_VSA_NEG	0.019(0.000)1	DASA	-0.026(0.002)5
PEOE_VSA_POS	-0.036(0.000)1	DCASA	-0.006(0.001)4
Q_VSA_FHYD	14.697(0.588)2	FASA_H	13.314(0.000)1
Q_VSA_HYD	0.022(0.000)1	std_dim1	-2.055(0.129)2
Q_VSA_NEG	0.015(0.000)1	std_dim2	3.901(0.014)2
Q_VSA_POL	-0.035(0.000)1	std_dim3	-8.753(0.000)1
vsa_hyd	0.070(0.018)10	x3D	
TPSA	-0.078(0.000)1	dipoleY	-4.419(0.318)2
logP (O/W)	-2.252(0.000)1	pmiY	-0.003(0.000)2

^a, the descriptors are identified from the one two and three parameter models emerging from CP-MLR with filter-1 as 0.74; filter-2 as 2.0; filter-3 as 0.74; filter-4 as $0.3 \leq Q^2 \leq 1.0$; **2D**: petitjean, (diameter - radius) / diameter; a_nCl, Number of chloro atoms; VSA(van der Waals surface area) descriptors with prefixes PEOE (partial equalization of orbital electronegativities) and Q (partial charges on the atoms of the structure): PEOE_VSA+2 and PEOE_VSA-1 (total VSA with partial atomic charges in the range (0.10,0.15) and (-0.10,-0.05) respectively; NEG, total negative VSA; POS, total positive VSA; FHYD, fractional hydrophobic VSA/ total surface area ($|q_i| \leq 0.2$); HYD, total hydrophobic VSA ($|q_i| \leq 0.2$); POL, total polar VSA ($|q_i| \geq 0.2$); vsa_hyd, total of VSA of hydrophobic atoms; TPSA, total polar surface area; logP (O/W), Log of the octanol/water partition coefficient; I_1 , when $R'' = \text{Me/H}$ then 1 otherwise zero (see ref. 10); **i3D**: rgyr, Radius of gyration; ASA (water accessible surface area) descriptors: ASA+, ASA of all atoms with positive partial charge; ASA_P, Water ASA of all polar atoms ($|q_i| \geq 0.2$); DASA, [(ASA+) - (ASA-)]; DCASA, [(Charge weighed ASA+) - (Charge weighed ASA-)]; FASA_H, (ASA_H / ASA); std_dim (Standard dimension 1,2&3): the square root of the 1st, 2nd & 3rd largest eigenvalue of the covariance matrix of the atomic coordinates; **x3D**: dipoleY, the y component of the dipole moment; pmiY, y component of the principal moment of inertia; also see ref 15.

^b, the average regression coefficient of the descriptor corresponding to all model, its standard deviation (s.d.) and the total number of its incidence. The arithmetic sign represents the actual sign of the regression coefficient in the models.

Table 5. Training and test sets statistics

Training equation's origin ^a	Set ^b	r	Q ²	test set r ^{2c}	s	F
Eq.2	1	0.867	0.558	0.564	0.625	10.109
	2	0.882	0.633	0.534	0.544	11.632
Eq.3	1	0.877	0.610	0.617	0.603	11.125
	2	0.873	0.589	0.692	0.562	10.648
Eq.4	1	0.924	0.520	0.778	0.506	13.124
	2	0.927	0.560	0.776	0.456	13.690
Eq.5	1	0.931	0.523	0.570	0.483	14.619
	2	0.950	0.732	0.527	0.379	20.855
Eq.6	1	0.888	0.652	0.546	0.577	12.426
	2	0.888	0.684	0.606	0.530	12.443
Eq.7	1	0.822	0.231	0.508	0.587	5.765
	2	0.789	0.102	0.651	0.653	4.555

^a, training and test set compounds are identified in Tables 1 and 2; number of compounds in training equations corresponding to Eqs.2 to 6 is fourteen and that of corresponding to Eq.7 is fifty;

^b, 1 for MACCS set and 2 for random set;

^c, in computing the test set r² values the mean observed activity of the training set has been used; for all the cases corresponding to Eqs 2 to 6 the test set size is seven compounds, and for the cases corresponding to Eq 7 it is twenty-five compounds.

Table 6. Summary of physicochemical meaning of identified descriptors (**Table 4**) in relation to the activity.

Descriptor	Inference from regression coefficients
petitjean; rgyr; pmiY	Bulky substitution at the periphery of molecular scaffold is undesirable.
Q_VSA_FHYD; Q_VSA_HYD; vsa_hyd; FASA_H	Hydrophobic surface areas of the compounds are desirable.
PEOE_VSA+2; PEOE_VSA_NEG; PEOE_VSA_FNEG; PEOE_VSA_POS; Q_VSA_NEG; Q_VSA_POL	Negatively charged / polarised surface areas are desirable.
ASA+; DCASA; DASA; ASA_P	Partially positively charged / polarized atomic regions are undesirable.
logP O/W	Overall hydrophobicity
Std_dim 1; Std_dim 3	Molecules with compact and closely arranged atoms are desirable.
Dipole	Molecular polarity is desirable.
Dipole Y	Directionality of dipole component for interaction with the receptor.
A_nCl ; I ₂	2,6-Dichloro substitution on 2-phenyl is desirable.
I ₁	Apolar environment in the vicinity of 3-pyrimidinyl moiety is desirable.