Combinatorial Chemistry and Highthroughput Screening (2005), 8, 431

Topological Descriptors in Modeling the HIV Inhibitory Activity of 2-Aryl-3-pyridyl-thiazolidin-4-ones

Y. S. Prabhakar, R. K. Rawal, M. K. Gupta, V. Raja Solomon and S. B. Katti*

Division of Medicinal and Process Chemistry,

Central Drug Research Institute, Lucknow – 226 001(U.P), India

Abstract

The HIV-1 RT inhibitory activity of 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones has been analyzed with different topological descriptors obtained from DRAGON software. Here, simple topological descriptors (**TOPO**), Galvez topological charge indices (**GVZ**) and 2D autocorrelation descriptors (**2DAUTO**) have been found to yield good predictive models for the activity of these compounds. The correlations obtained from the **TOPO** class descriptors suggest that less extended or compact saturated structural templates would be better for the activity. The participating **GVZ** class descriptors suggest that they have same degree of influence on the activity. In **2DAUTO** class, the large participation of descriptors of lags seven and three indicate the association of activity information with the seven and three centered structural fragments of these compounds. The physicochemical weighting components of these descriptors suggest homogeneous influence of mass, volume, electronegativity and/ or polarizability on the activity.

Keywords: 2-(2,6-Dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones, HIV-1 RT, combinatorial protocol in multiple linear regressions (CP-MLR), DRAGON's topological descriptors, QSAR study.

Introduction:

The Reverse Transcriptase (RT) of Human immunodeficiency virus type-1 (HIV-1) is the prime target for the development of drugs for HIV/AIDS therapy [1]. Here the non-nucleoside class of compounds commonly addressed as non-nucleoside RT inhibitors or NNRTIs, bind allostrically to hydrophobic pocket close to active site and achieve highly selective suppression of HIV-1 replication with little cytotoxicity to the host cells and have become inhibitor of choice for HIV-1 RT [2,3]. Because of this considerable research efforts have been focused on the synthesis and structure-activity relationships of a large numbers of different chemical scaffolds as NNRTIs [4]. Also, the X-ray crystallographic studies of diverse NNRTIs/RT complexes have shown that the NNRTIs, irrespective of their structure class, uniformly assume "butterfly-like" shape in the

[#] CDRI Communication no.6585

* Corresponding author. Fax: +91-522-223405, e-mail: setu_katti@yahoo.com

inhibition of HIV-1 RT [5]. In this background a quantitative structure-activity relationship (QSAR) study of the HIV-1 RT inhibitory activity of two series of NNRTI class compounds, 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones [6] and 2-(2,6-Dihalophenyl)-3-(substituted phenyl)-thiazolidin-4-ones [7] belonging to 2,3-diaryl-thiazolidin-4-ones, has been carried out in terms of physicochemical and structural descriptors [8]. This study too suggested that compounds having the ability to take butterfly like conformation would be showing better inhibitory activity. Also, it indicated a preference for hydrophobic compounds and the ability of the meta-positions of 3-aryl moieties to accommodate hydrophobic/ steric groups. Furthermore the developed models suggested that the factors responsible for the variation in the activity have been uniformly distributed across the varying centers of the molecule. Even though from geometry point of view, these two series of compounds span almost the same structural space and steric features, presence of nitrogen in the 3-aryl moiety of 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones gave them a blend of polarity, electronic features and a different hydrophobicity profile when compared to corresponding 3-phenyl analogue. These characteristic features of 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones have prompted us to further investigate the structure-activity relations of these compounds with a wider dragnet of descriptors. In this connection, application of graph theory to chemical structure opened the ways for the computation of several hundreds of topological, topographical and related descriptors to characterize the molecule from different perspectives. DRAGON [9], TOPS-MODE [10], MOLCONN [11], CODESSA [12], POLLY [13] are some well known or recent programs embedded with graph theoretical concepts for characterizing the chemical structure and offer large number of descriptors for modeling studies. When dealing with such a large number of descriptors in the model development, for the optimum utilization of contents of the generated datasets it is necessary to haul out the best models as well as information rich descriptors corresponding to the phenomenon under investigation. The Genetic Function Approximation (GFA) [14], MUtation and SElection Uncover Models (MUSEUM)[15] and Combinatorial Protocol in Multiple Linear Regression (CP-MLR)[16] are a few approaches that address the evolution of multiple models and in so doing identify contributing descriptors in quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) studies. Here we have considered the CP-MLR approach [8,16,17] to discover a wide range of structure-activity models and contributing descriptors for the HIV-1 RT inhibitory activity of 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Figure-1) [6] in terms of different empirical, constitutional and graph theoretical descriptors opted from DRAGON software [9]. In this, each model may address different sub-structural regions and attributes in the predictive and diagnostic aspects of the chosen phenomenon. Here, the multi-descriptor class environment offers scope to explore, analyze and understand the phenomenon under investigation vis-à-vis the descriptors. The populations of the descriptors of the models provide scope to understand the predictive and diagnostic aspects of different sub-structural regions and in averaging their influence beyond the individual models. With this perspective we contemplated a QSAR study on the 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (figure-1) [6] to explore the different structural attributes and their information content in rationalizing the activity of these analogues. The results are presented here.

2 Materials and Methods

2.1 Dataset

The chemical structure information of 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones and their reported anti-HIV activity (EC_{50}, i.e., concentration, in moles per liter, required to reduce HIV-1 induced cytopathic effect by 50% in MT-4 cells in the form of logarithm of the inverse of inhibitory concentration and expressed as $-\log EC_{50}$) have been listed in Table 1. For the computation of descriptors in DRAGON program [9], the appropriate structure files of the compounds have been prepared in CS Chem3D Ultra [18] and HyperChem [19]. The descriptor classes considered in the study along with their definitions and scopes in addressing the molecular structure have been presented in Table 2. As the total number of descriptors involved in this study is two hundred and seventy-seven, only the names of descriptor classes and participating descriptor names have been used in the discussion [20]. The CP-MLR computational procedure [16] and model validation are briefly described below.

2.2 Computational Procedure

CP-MLR is a 'filter' based variable selection procedure for model development in QSAR and QSPR studies. It involves selected subset regressions [16]. 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, each having unique combination of descriptors from the dataset under study. In this, the contents and number of variables to be evaluated are mixed according to the predefined confines. Here the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables [21]. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.74) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms 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 addition to cross-validation, each identified model has been reassessed for the chance correlations, if any, by repeated randomization of the biological response [8, 22]. The datasets with randomized response vector have been subjected to multiple regression analysis. The emerging regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. Every model has been subjected to one hundred such simulation runs. This has been used as a measure to express the percent chance correlation of the model under examination.

Results and Discussion:

Exploring for models along the descriptor class provides an opportunity to assess the information content of each class with respect to the phenomenon under investigation. In this study, unless otherwise stated, the four 'filters' embedded in CP-MLR have been set to their default values, that is filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.74 and filter-4 as $0.3 \leq Q^2 \leq 1.0$. With this, attempts have been made to develop one- and two-descriptor models within each class of descriptors for the HIV-1 RT inhibitory activity of 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Table-1). The filter-3 with 0.74 will allow the collection of all models with explained variance beyond 54 per cent and makes it possible to identify all descriptors having the information content to explain the phenomenon under investigation. These study conditions have led to thirty-one models in the topological (**TOPO**) class descriptors, thirty-four models in Galvez Topological charge indices (**GVZ**) and one hundred and forty-four models in the 2D autocorrelations (**2DAUTO**) class descriptors (Table-2). The essence of all the models obtained from the study has been provided in Table-3 in the form of identified descriptor's average of regression coefficients along with standard deviation across the models and the total incidence corresponding to all the models. This, while providing the averages of the estimated regression coefficients of all the identified descriptors, shows their variance across the models emerging from the study as well. To maintain brevity, the complete regression equations have been shown for selected models only. The following regression equations represent **TOPO** class structure-activity model of the compounds (Table- 1).

$$-\text{LogEC}_{50} = 126.245 - 1.874(0.228)\text{AECC} - 300.003(30.650)\text{PW3}$$
$$n=24,\ r=0.941,\ Q^2=0.855,\ s=0.308,\ F=80.89 \quad (1)$$

$$-\text{LogEC}_{50} = 134.429 - 5.866(0.752)\text{ICR} - 324.197\ (31.991)\text{PW3}$$
$$n=24,\ r=0.936,\ Q^2=0.834,\ s=0.320,\ F=74.32 \quad (2)$$

In this and 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. Also, in the randomization study (hundred simulations per model) none of the identified models has shown any chance correlation. The thirty-one models, including the above-presented two, emerged from **TOPO** class have shared seventeen descriptor among themselves (Table 3). Among these descriptors, path/walk-3 ratio (PW3) has the highest frequency of eight, which is also part of equations 1 and 2. All the path/walk indices are Randic's molecular shape descriptors and their values increase with increased branching in the vertices. Of all the **TOPO** class descriptors PW3 alone correlated with the activity to the order of 0.72. The other path/walk indices contributed to the model formation are PW4 (frequency: six) and PW5 (frequency: three) (Table-3). The other descriptor took part in equation 1 corresponds to average eccentricity (AECC; frequency: three) of the compound. This index represents the eccentric nature of the topological graph. The other eccentric or centric related descriptors participated in the

study are radial centric information index (ICR; frequency: four) in equation 2, total eccentricity (TECC; frequency: one) and eccentric connectivity index (CSI; frequency: three), mean information content on the distance degree equality (IDDE; frequency: six), Petitjean 2D shape index (PJI2; frequency: five) and variation (VAR; frequency: one). All these descriptors are highly inter-correlated ($r > 0.8$) and are associated with negative regression coefficient. Five information content descriptors (IC, TIC, SIC, CIC and BIC) corresponding to neighborhood symmetry are among the other contributing descriptors (Table-3) in modeling the HIV-1 RT inhibitory activity of the compounds. All these five IC descriptors are inter-correlated to different degree (r is 0.62 to 0.98). Only one Kier and Hall's molecular connectivity index in the form of average connectivity index of order-2 (X2A) has participated in five models. The regression coefficients associated with the identified **TOPO** descriptors (Table 3) suggest that compounds with a less extended topological graph would be better for the HIV-1 RT inhibitory activity.

The thirty-four models resulted from Galvez topological charge indices (**GVZ**) have shared eighteen descriptors among themselves (Table-3). In this class, all descriptors, baring three, contributed to the modeling of HIV-1 RT inhibitory activity of the compounds (Table-3). These descriptors have their origin in the first ten eigenvalues of the polynomial of corrected adjacency matrix of the compounds. All the **GVZ** class descriptors belong to two categories. Of this one category corresponds to the topological charge index of order n (GGIn) and the other to the mean topological charge index of order n (JGIn), where 'n' represents the order of eigenvalue. Interestingly for both GGIn and JGIn, all eigenvalues, excepting the ninth one, found place in one or other model and among these descriptors, the one corresponding to seventh eigenvalue has the highest incidence (table-3). The magnitude of regression coefficient of both the odd and even eigenvalues, increased with increasing order of the eigenvalue. As the eigenvalues represent the characteristic roots of a polynomial arranged in descending order, this may suggest that, irrespective of their order, all eigenvalues may have been embedded with comparable level of information in them. In all the two-descriptor models, one participating descriptor has come from even order eigenvalue associated with negative regression coefficient and the other descriptor has come from odd order eigenvalue associated with positive regression coefficient. Furthermore, an examination of the correlation matrix of the identified **GVZ** descriptors indicated that for the same order of eigenvalues, GGIn and JGIn are highly correlated (r is 0.9 to 1.0). All these suggest a clear parallelism between GGIn and JGIn descriptor groups. The frequencies of identified descriptors clearly indicate that descriptors from the mean topological charge index (JGIn) have contributed to more models when compared to that of the ordinary topological charge index (GGIn).

$$-\text{LogEC}_{50} = 21.831 - 263.490(24.496)\text{JGI4} + 8.236(0.923)\text{GGI7}$$
$$n=24,\ r=0.944,\quad Q^2=0.859,\quad s=0.301,\quad F=85.39 \qquad (3)$$

$$-\text{LogEC}_{50} = 22.699 - 278.182(25.905)\text{JGI4} + 174.363(20.664)\text{JGI7}$$
$$n=24,\ r=0.938,\quad Q^2=0.834,\quad s=0.314,\quad F=77.42 \qquad (4)$$

The above two equations sharing GGI7, JGI4 and JGI7 (equations 3 and 4) represent the best models emerging from **GVZ** class descriptors. As GGI7 and JGI7 are highly inter-correlated (r=0.994), these equations may be treated as replica of one another.

The **2DAUTO** class has contributed maximum number of descriptors (sixty-one) for modeling the HIV-1 RT inhibitory activity of 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Tables-1 and-3). These descriptors also represent the topological structure of the compounds. But they are more complex in nature when compared to the **TOPO** class or even **GVZ** class of descriptors. The **2DAUTO** descriptors considered in the study have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS) [9]. The computation of these descriptors involve the summations of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments [23]. Also a weighting component in terms of a physicochemical property has been embedded in this descriptor. As a result these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag n (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component – m for mass or v for volume or e for Sanderson electronegativity or p for polarizability – for its computation. The philosophy embedded in these descriptors and their computational aspects are available in different sources [9,23]. The following two regression equations (equations 5 and 6) have been selected to represent the models of this class of descriptors.

$$-\text{LogEC}_{50} = 9.191 + 4.462(1.314)\text{MATS3v} - 1.037(0.103)\text{GATS8e}$$
$$n=24,\ r=0.931,\ Q^2=0.824,\ s=0.331,\ F=68.59 \qquad (5)$$

$$-\text{LogEC}_{50} = 12.430 - 3.475(0.392)\text{MATS7m} - 3.265(0.621)\text{GATS3v}$$
$$n=24,\ r=0.924,\ Q^2=0.808,\ s=0.346,\ F=61.70 \qquad (6)$$

In these two equations Geary autocorrelation lag eight weighed by electronegativity (GATS8e) and Moran autocorrelation lag seven weighed by atomic masses (MATS7m) are correlated with the activity (r is 0.89 and 0.81, respectively). Apart from these two, five more descriptors (MATS5e, MATS6e, MATS7e, MATS8e and MATS8m) have shown direct correlation with the activity. Also, all these seven descriptors are highly inter-correlated (r>0.86). Furthermore, the trend of the formed models and descriptors participated therein indicate that descriptors of lag seven have accounted for the maximum number of models (cumulative frequency is eighty) followed by lag three (cumulative frequency is seventy-four) and lag four (cumulative frequency is forty-three) descriptors (Table-3). Also, among the autocorrelation descriptors, the ones from Moran (MATS) and Geary (GATS) have contributed to the maximum models. Moreover, the autocorrelation descriptors of Broto-Moreau (ATS) only contributed in association with Moran (MATS) and Geary (GATS) descriptors. A study of the correlation matrix of these

descriptors revealed that in most of the cases the autocorrelation vectors of each lag formed by volume and polarizability weightings are highly inter-correlated. Also the Moran (MATS) and Geary (GATS) descriptors of respective lags and weighting components are inter-correlated suggesting that these descriptors carry similar information content. The population of emerged models and the participating descriptors therein suggest that descriptor's physicochemical weighting components mass, volume, electronegativity and/ or polarizability have homogeneous influence on the activity. The large participation of descriptors of lag seven and three may be viewed in terms of association of activity information with the seven and three centered structural fragments of the compounds under study.

In the light of models obtained from **TOPO**, **GVZ** and **2DAUTO** classes, the structure-activity relations of these compounds (Table 1) have been reinvestigated by pooling all the descriptor of these three classes into one dataset. This resulted in many inter-descriptor class models with equally good or higher predictivity. The following equation is one such model from **TOPO** and **GVZ** class descriptors (equation 7).

$$-\text{LogEC50} = -98.904 + 388.330(40.234)\text{X2A} - 10.280(1.031)\text{GGI6}$$
$$n=24,\ r=0.946,\ Q^2=0.866,\ s=0.294,\ F=89.86 \qquad (7)$$

However, in the descriptor classes **EMP**, **CONST**, **MWC** and **BCUT**, which have not yielded any model individually, even the pooling of all descriptor into one dataset has not result in any model under the study conditions. Also, our attempts by combining the descriptors of each of the **TOPO**, **GVZ** and **2DAUTO** classes with **EMP**, **CONST**, **MWC** and **BCUT** classes also did not result in any better predictive model for the activity. This while suggesting the usefulness of **TOPO**, **GVZ** and **2DAUTO** class descriptors for modeling the HIV-1 RT inhibitory activity of 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Table 1), also suggests that the other descriptor classes namely **EMP**, **CONST**, **MWC** and **BCUT**, have insufficient or marginal information (corresponding to the activity) within themselves – independently as well as collectively – to come up to the level of model formation.

Conclusions:

In summary, the HIV-1 RT inhibitory activity of 2-(2,6-dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Table-1) a correlation has been found with simple topological (**TOPO**), Galvez topological charge indices (**GVZ**) and 2D autocorrelations (**2DAUTO**) descriptors of these compounds. The activities predicted by the selected models of this study are in agreement with the observed ones (Tables-1). The descriptors from other classes, that is, empirical (**EMP**), constitutional (**CONST**), molecular walk counts (**MWC**) and modified burden eigenvalues (**BCUT**) have little or insignificant information corresponding to the activity. The correlations obtained from the **TOPO** descriptors suggest that less extended or compact saturated structural templates would be better for the activity. The participating **GVZ** descriptors and their regression coefficients indicate that all eigenvalues have comparable influence on the activity. In **2DAUTO** class, the trend of the formed models and descriptors participated therein indicate that descriptors of lag seven have accounted for the maximum number of models followed by

lags three and four (Table-3). Furthermore, the study suggests that the descriptor's physicochemical weighting components mass, volume, electronegativity and/ or polarizability have homogeneous influence on the activity. The large participation of descriptors of lag seven and three suggests the association of activity information with the seven and three centered structural fragments of the compounds under study.

References:

[1] Clercq, E.De. *Rev. Med. Virol.* **2000**, *10,* 255.
[2] Clercq, E.De. *Med.Res.Rev.* **1993**, *13*, 229.
[3] Hajos, G.; Riedi, S.; Monar, J.; Szabo, D. *Drugs Fut.* **2000***, 25*, 47.
[4] Garg, R.; Gupta, S.P.; Gao, H.; Babu, M.S.; Debnath, A.K.; Hansch, C. *Chem.Rev.* **1999**, *99*, 3525.
[5] Schafer, W.; Friebe, W.G.; Leinert, M.; Merttens, A.; Poll, T.; Von der Saal, W.; Zilch, H.; Nuber, H.; Ziegler, M.L. *J. Med. Chem.* **1993**, *36*, 726.
[6] Barreca, M.L.; Balzarini, J.; Chimirri, A.; Clercq, E. De.; Luca, L. D; Holtje, H.D.; Holtje, M.; Monforte, A.M.; Monforte, P.; Pannecouqe, C.; Rao, A.; Zapalla, M. *J. Med. Chem*. **2002**, *45*, 5410.
[7] Rao, A.; Carbone, A.; Chimirri, A.; Clercq, E. D.; Monforte, A.M.; Monforte, P.; Pannecouque, C.; Zappala, M. *Il Farmaco*. **2002**, *58*, 115.
[8] Prabhakar, Y.S.; Solomon, V.R.; Rawal, R.K.; Gupta, M.K.; Katti, S.B. *QSAR and Combi. Sc.* **2004**, *23*, (In press).
[9] Dragon software (version **1.11-2001**) by Todeschini R.; Consonni V. Milano, Italy and references cited therein, http://disat.unimib.it/chm/Dragon.htm
[10] Gonzalez, M.P.; Helguera, A.M. *J. Comput.-Aided Mol. Des.*, **2003**, *17*, 665.
[11] Hall, L.H.; MOLCONN-X; Hall Associates Consulting, Quincy: MA, **1991**.
[12] (a) Katritzky, A.R; Lobnov, V; Karelson, M. *CODESSA* (*Comprehensive Descriptors for Structural and Statistical Analysis*) University of Florida: Gainesville, FL, 1994; (b) Katritzky, A.R.; Perumal, S.; Petrukhin, R.; Kleinpeter, E.; *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 569.
[13] Basak, S.C. POLLY; Natural Resource Research Institute, Duluth, University of Minnesota: Duluth, MN, **1988**.
[14] Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
[15] Kubinyi, H. *Quant. Struct.-Act. Relat*. **1994**, *13*, 285.
[16] Prabhakar, Y. S. *QSAR and Combi. Sc.* **2003**, *22*, 583.
[17] Prabhakar, Y. S. *Internet Electron. J. Mol. Des.* **2004**, *3*, 150. http://www.biochempress.com
[18] ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA.
[19] HyperChem Release 3, Autodesk Inc., California USA, **1993**.
[20] The complete descriptor dataset of all compounds will be provided on request.
[21] Armitage, P.; Berry, G. *Statistical Methods in Medical Research. 2nd Edition*", Blackwell Scientific Publications, Oxford, **1990**, pp.296-357.
[22] So, S.S.; Karplus, M.; *J. Med. Chem*. **1997**, *40*, 4373.
[23] Broto, P.; Moreau, G.; Vandycke, C.; *Eur. J. Med. Chem*. **1984**, *19*, 66.

Table-1: 2-(2,6-Dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (Figure-1) and their HIV-1 RT inhibitory activity.

Comp	R″				R′		$-\log EC_{50}$				
No.	3″	4″	5″	6″	2′	6′	Obsd[a]	Eq.1	Eq.3	Eq.5	Eq.7
1	H	H	H	H	Cl	Cl	6.75	6.53	6.50	6.86	6.45
2	H	H	H	H	Cl	F	6.56	6.53	6.50	6.41	6.45
3	H	H	H	H	F	F	6.07	6.53	6.50	6.06	6.45
4	H	H	Cl	H	Cl	Cl	5.75	5.59	5.65	5.89	5.57
5	H	H	Cl	H	Cl	F	5.67	5.59	5.65	5.64	5.57
6	H	H	Cl	H	F	F	5.14	5.59	5.65	5.40	5.57
7	H	H	Br	H	Cl	Cl	5.82	5.59	5.65	5.97	5.57
8	H	H	Br	H	Cl	F	5.91	5.59	5.65	5.68	5.57
9	H	H	Br	H	F	F	5.31	5.59	5.65	5.44	5.57
10	H	H	H	Br	Cl	Cl	6.56	6.96	7.00	7.35	7.05
11	H	H	H	Br	Cl	F	7.19	6.96	7.00	6.99	7.05
12	H	H	H	Br	F	F	7.52	6.96	7.00	6.71	7.05
13[b]	CH_3	H	H	H	Cl	Cl	-	4.35	4.05	5.49	4.38
14	CH_3	H	H	H	F	F	4.27	4.35	4.05	4.25	4.38
15	H	CH_3	H	H	Cl	Cl	6.83	6.96	7.00	7.09	7.05
16	H	CH_3	H	H	Cl	F	7.00	6.96	7.00	6.74	7.05
17	H	CH_3	H	H	F	F	6.60	6.96	7.00	6.46	7.05
18	H	H	CH_3	H	Cl	Cl	5.85	5.58	5.65	5.83	5.57
19	H	H	CH_3	H	F	F	5.29	5.59	5.65	5.72	5.57
20	H	H	H	CH_3	Cl	Cl	7.36	6.96	7.00	7.47	7.05
21	H	H	H	CH_3	Cl	F	7.28	6.96	7.00	7.10	7.05
22	H	H	H	CH_3	F	F	7.09	6.96	7.00	6.81	7.05
23	H	CH_3	H	CH_3	Cl	Cl	7.04	7.40	7.24	7.48	7.25
24	H	CH_3	H	CH_3	Cl	F	7.38	7.40	7.24	7.18	7.25
25	H	CH_3	H	CH_3	F	F	7.23	7.40	7.24	6.94	7.25

[a], ref. 6
[b], not included in the analysis

Table-2: The descriptors classes used for the analysis of 2-(2,6-Dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (**Figure-1; Table-1**) and number of models identified in each class.

Descriptor Class (acronym)[a]	Definition and scope	Contributing descriptors	Descriptors per model (Number of models)[b]
Empirical (**EMP**)	Represents the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in H-depleted molecule		
Constitutional (**CONST**)	Dimensionless descriptors, independent from molecular connectivity and conformations		
Topological (**TOPO**)	Descriptors from topological graphs and independent of conformations	17	2(31)
Molecular Walk counts (**MWC**)	Descriptors representing self returning walks counts of different lengths		
Modified Burden eigenvalues (**BCUT**)	Descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements & atoms		
Galvez Topological charge indices (**GVZ**)	Descriptors representing the first 10 eigenvalues of corrected adjacency matrix	18	1(1); 2(33)
2D auto correlations (**2DAUTO**)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)	61	1(7); 2(137)

[a], ref.9

[b], models emerged from CP-MLR protocol with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.74; filter-4 as $0.3 \leq Q^2 \leq 1.0$, number of compounds in the study are twenty-four.

Table 3: Descriptors identified in modeling the HIV-1 RT inhibitory activity of 2-(2,6-Dihalophenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones (**Table 1**) along with their average regression coefficients

Descriptor[a]	Av Reg Coef (sd) Total incidence[b]
TOPO	
X2A	348.772(36.507)5
ICR	-4.572(1.223)4
IDDE	2.541(0.262)6
IC	5.530(0.486)4
TICE	0.139(0.040)3
SIC	27.064(2.308)2
CIC	-6.087(0.444)2
BIC	28.796(2.550)3
S3K	2.104(--)1
CSI	-0.029(0.008)3
PJI2	-5.910(0.969)5
TECC	-0.068(--)1
AECC	-1.790(0.097)3
VAR	-0.088(0.017)3
PW3	-309.704(39.558)8
PW4	106.977(17.250)6
PW5	-459.206(76.956)3
GVZ	
GGI1	0.920(0.025)2
GGI2	-3.907(0.446)3
GGI3	2.926(0.506)4
GGI4	-6.571(--)1
GGI5	7.697(--)1
GGI6	-10.226(1.804)4
GGI7	6.841(1.224)8
GGI8	-14.954(1.426)3
GGI10	-90.623(4.704)2
JGI1	17.536(6.861)3
JGI2	-100.256(3.998)4
JGI3	101.104(20.881)3
JGI4	-253.514(26.416)4
JGI5	231.818(91.626)3
JGI6	-229.062(23.766)7
JGI7	137.946(36.068)9
JGI8	-280.903(24.777)4
JGI10	-1087.473(56.444)2
2DAUTO	
ATS6m	1.278(0.103)2
ATS7m	-1.295(0.314)2
ATS8m	0.440(--)1
ATS2v	-9.619(--)1
ATS3v	8.477(--)1
ATS6v	8.997(0.618)3
ATS8v	4.992(--)1
ATS1e	-27.853(--)1
ATS3e	-21.440(--)1
ATS7e	-15.515(--)1
ATS8e	-14.458(2.411)7
ATS3p	5.765(--)1
ATS5p	6.047(--)1
ATS6p	6.337(0.360)3
ATS8p	3.314(--)1
MATS2m	9.232(1.794)5
MATS3m	6.862(--)1
MATS4m	2.510(--)1
MATS5m	-11.791(0.866)2
MATS6m	5.021(0.955)8
MATS7m	-3.895(0.263)17
MATS8m	2.623(--)1
MATS1v	112.637(--)1
MATS3v	7.657(1.943)14
MATS4v	-9.728(3.098)10
MATS5v	6.228(1.800)2
MATS6v	-3.462(0.576)2
MATS7v	-5.010(0.751)5
MATS1e	48.944(--)1
MATS2e	17.323(1.302)2
MATS3e	-18.517(4.136)6
MATS4e	10.025(1.731)7
MATS5e	-14.286(0.721)3
MATS6e	14.746(0.707)3
MATS7e	-8.637(0.385)4
MATS8e	4.230(--)1
MATS3p	9.480(0.864)13
MATS4p	-6.689(2.791)8
MATS5p	7.152(1.842)2
MATS7p	-3.856(0.667)8
GATS2m	-1.040(--)1
GATS3m	-2.300(0.884)4

Table-3 contd…

Table-3 contd

Descriptor[a]	Av Reg Coef (sd) Total incidence[b]	Descriptor[a]	Av Reg Coef (sd) Total incidence[b]
GATS5m	0.977(0.215)2	GATS4e	-3.16(0.465)4
GATS7m	1.369(0.127)17	GATS5e	3.751(0.222)4
GATS8m	-0.578(0.084)7	GATS6e	-2.469(0.072)2
GATS3v	-4.056(0.513)12	GATS7e	2.584(0.098)3
GATS4v	6.993(0.786)7	GATS8e	-1.076(0.035)4
GATS5v	-4.080(--)1	GATS3p	-4.426(0.632)16
GATS6v	1.878(0.168)3	GATS4p	4.052(1.025)6
GATS7v	5.320(0.539)12	GATS5p	-4.085(--)1
GATS2e	-8.932(1.892)5	GATS7p	1.866(0.172)11
GATS3e	3.818(0.791)5		

[a], the descriptors are identified from the one and two parameter models emerging from CP-MLR with filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.74; filter-4 as $0.3 \leq Q^2 \leq 1.0$.; **TOPO:** X2A, average connectivity index chi-2; ICR, radial centric information index; S3K-3, path Kier shape index; CSI, eccentric connectivity index; PJI2, 2D Petitjean shape index; TECC, Total eccntricity; AECC, Average eccentricity; VAR, variation; PWn, path/ walk n (Rendic shape) **neighborhood symmetry:**IC, information content index ; TIC, total IC; SIC, structural IC; CIC, complementary IC; BIC, bond IC; IDDE, mean IC on the distance degree equality;**GVZ:** topological charge index of order n (GGIn) and mean topological charge index of order n (JGIn), where 'n' represents the order of eigenvalue; **2DAUTO:** In the descriptors of autocorrelation of topological structure of ATSnw (Broto-Moreau), MATSnw (Moran) and GATSnw (Geary) the penultimate character, 'n' indicates the autocorrelation vector of lag n corresponding to the number of edges in the fragment unit considered in the computation and the last character 'w' refers to the weighting property used in the computation which may be either 'm' i.e. atomic masses or 'v' i.e. atomic van der Walls volumes or 'e' i.e. Sanderson electro negativities or 'p' i.e. atomic polarizabilities. See ref.9

[b], the average regression coefficient of the descriptor corresponding to all models, 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.

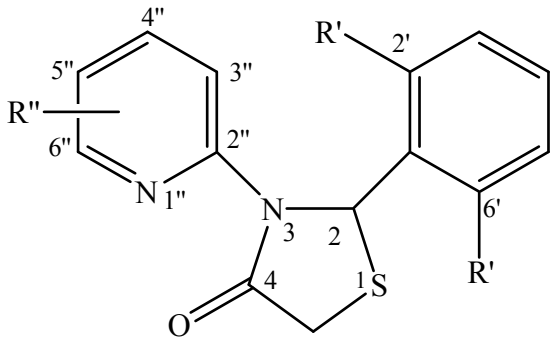

Figure 1

Legend for figure

Figure-1: 2-(2,6-dihalo phenyl)-3-(substituted pyridin-2-yl)-thiazolidin-4-ones