Central European Science Journals
w w w . c e s j . c o m

Central European Journal of Chemistry
Central European Science Journals
DOI: 10.1007/s11532-005-0010-0
Research article
CEJC 4(1) 2006 135–148

Improved QSAR modeling of anti-hiv-1 acivities by means of the optimized correlation weights of local graph invariants

Damián José Gabriel Marino[1], Eduardo Alberto Castro[2*], Andrey Toropov[3]

[1] *CIMA, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calles 47 y 115, La Plata 1900, Buenos Aires, Argentina*

[2] *INIFTA, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Suc.4, C.C. 16, La Plata 1900, Argentina*

[3] *Scientifical Research Institute "Algorithm – Engineering", F. Khodjaev Street 25, 700125 Tashkent, Uzbekistan*

Received 23 May 2005; accepted 9 August 2005

Abstract: We report the results derived from the use of molecular descriptors calculated with the correlation weights (CWs) of local graph invariants for modeling of anti-HIV-1 potencies of two groups of reverse transcriptase inhibitors. The presence of different chemical elements in the molecular structure of the inhibitors and the Morgan extended connectivity values of zeroth-, first-, and second order have been examined as local graph invariants in the labeled hydrogen–filled graphs. We have computed via Monte Carlo optimization procedure the values of CWs which produce the largest possible correlation coefficient between the numerical data on the anti-HIV-1 potencies and those values of the descriptors on the training set. The model of the anti-HIV-1 activity obtained with compounds of training set by means of optimization of correlation weights of chemical elements present together with Morgan extended connectivity of first order makes up a sensible model for a satisfactory prediction of the endpoints of the compounds belonging to the test set.

Keywords: QSAR Modeling – Anti-HIV-1 Activity – Correlation weight of local graph invariants – Flexible topological descriptors

* E-mail: castro@quimica.unlp.edu.ar; jubert@arnet.com.ar

1 Introduction

Human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2, respectively) are responsible for the condition known as AIDS (acquired immunodeficiency syndrome) and they differ in nucleotide and amino acid sequences [1, 2]. However, HIV-1 is the predominant type and is a retrovirus, i.e. it is an RNA virus that utilizes an enzyme known as DNA polymerase or reverse transcriptase (RT) to produce a DNA provirus that is capable of inserting into the host DNA. Since a key HIV replication step is the reverse transcription of genomic RNA into double stranded DNA, a widely used therapeutic strategy is to inhibit the RT enzyme that promotes this process [3, 4].

Since 1991, compounds possessing anti-HIV activity have been the subject of numerous QSAR studies. A search in the Chemical Abstract database [5] for the "QSAR and HIV" query led to about 200 references describing structure-property relationships established using partial least-squares (PLS), artificial neural network (ANN), and multiple linear regression (MLR) methods involving one dimensional (1D) and/or two-dimensional (2D) descriptors [6–11] or three-dimensional (3D) descriptors [12, 13], the four-dimensional (4D)-QSAR technique [14], comparative molecular field analysis (CoMFA) [15–19] and electrostatic potential distribution [20, 21].

Quantitative Structure–Activity Relationships (QSAR) modeling is a statistical analysis, developed by Hansch, to elucidate a quantitative correlation between chemical structure and biological activity. QSAR methods have been widely applied to many drug design problems. The fundamental hypothesis of QSAR is that biological properties are functions of molecular structure. Developments in modeling the relationship between molecular structure and various drug properties have taken two distinct routes. Approaches that require significant information from 3D molecular geometries of both the enzyme active site as well as for drug molecules are often called rational or structure-based design methods. These methods require individual atom coordinates for the active site, minimum energy conformations for inhibitors, individual docking of each inhibitor into the active site, and detailed intermolecular interaction energy calculations. Consequently, these methods are time-consuming and expensive, although they have demonstrated quite good capabilities for useful drug design. The other line of approach is considerably less time-consuming and less expensive. The methods involved are based on topological representation of molecular structure and do not require 3D-based geometry information of the active site or of the drug molecules. In this approach, we employ topological superposition of the common core skeleton. Furthermore, the methods for computation are very fast, significantly faster than the methods that require 3D geometry information. The models produced in this topological approach also yield significant structure information for the design of new compounds [22–29].

Recently, the Optimization of Correlation Weights of Local Graph Invariants (OCWLGI) has been suggested as a suitable approach of the QSAR analyses [30–40]. This particular set of molecular descriptors belongs to the category that encompasses more flexible descriptors. (more encompassing category of flexible descriptors). The concept of flexi-

ble topological descriptors, originally introduced by Randic [41–43], is a major advance with regard to the possibility of extracting maximum amount of chemical information. At the same time, the descriptors used in the multiple regression equation are not intercorrelated among themselves. Since, the regression procedure uses a single descriptor the difficulties associated with multiple regression are not present in such an approach. Unlike usual fixed topological descriptors, flexible topological descriptors do not have a definite predetermined value, which can (which can or that can) be applied to any sets of compounds for the modeling of physicochemical property or/and biological activity. In fact, the formalism of such descriptors is defined on the basis of an optimization procedure to get the best optimal relation for a particular data set. Thus, the definition of the descriptors will vary between data sets and the ultimate purpose of the iterative optimization procedure is to obtain the best predictive model.

The aim of the present study is to apply an improved approach to the optimization of correlation weights scheme for modeling the anti-HIV-1 properties of inhibitors detailed by Gonzáles et. al. [22]. Two sets of data have been employed for the present analysis: TIBO derivatives and HEPT derivatives (Figure 1). In a series of studies Ho *et al.* demonstrated that certain compounds of type **1** show anti-HIV-1 activity by functioning as non-nucleoside reverse transcriptase (RT) inhibitors [44]. They systematically synthesized and tested numerous members of this family, differing in substituents X, Y, Z, and R. Here, Z was restricted to oxygen and/or sulfur. The potency was found to be enhanced by a 3,3-dimethylallyl substitution at the 8-position with Z = S. HEPT derivatives belong to type **2** class of HIV-1 RT inhibitors [6, 45]. The X can be oxygen or sulfur however, it was observed that the former tends to produce a higher level of activity.

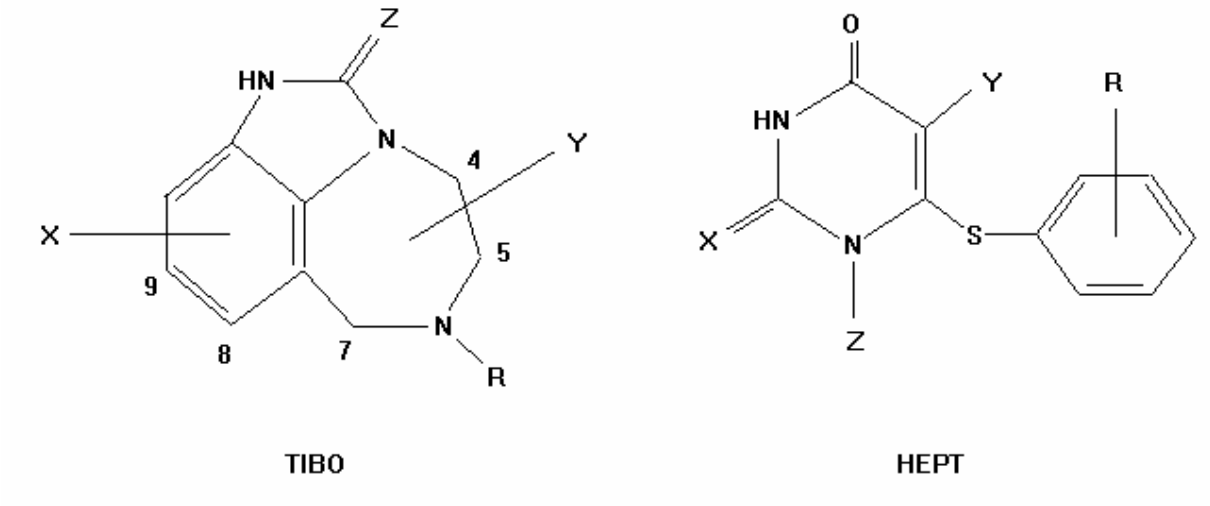

Fig. 1 The TIBO and the HEPT derivatives, differing in X,Y,Z, and R substituents.

Effectiveness in inhibiting HIV-1 was measured by the concentration of the compound, C_{50}, required to achieve 50 % protection of MT-4 cells against the virus [6, 44, 45]. We considered the experimental results for thirty eight TIBO and nineteen HEPT derivatives as representative for developing regression equations for anti-HIV-1 potencies of these

classes of compounds.

2 Method

Models of anti-HIV-1 activities which have been examined in the present study are based on labeled hydrogen filled graph (LHFG) and described by equation 1:

$$DCW(a,^{x}EC) = \sum_{k=1}^{n}[CW(a_k) + CW(^{x}EC)] \quad (1)$$

In the above equation, the DCW term represents the molecular descriptor, CW terms stand for the correlation weights, a_k is the chemical element which is image of the kth vertex of the LHFG, and ^{x}EC is the Morgan extended connectivity [13]. As local invariants, we have used the Morgan extended connectivity of zero, first, and second order (denoted by ^{0}EC, ^{1}EC, and ^{2}EC, respectively).

Zero-order Morgan connectivity of an atom k is the adjacency count of that atom. The first-order Morgan connectivity value of atom k is the sum of the zero-order Morgan connectivity values of the atoms that are connected to atom k. Similarly, the second-order Morgan connectivity value of atom $k$is the sum of the first-order Morgan connectivity values of the atoms that are connected to atom k. An example of calculations of Morgan connectivity indices for ethanol is given in Figure 2.

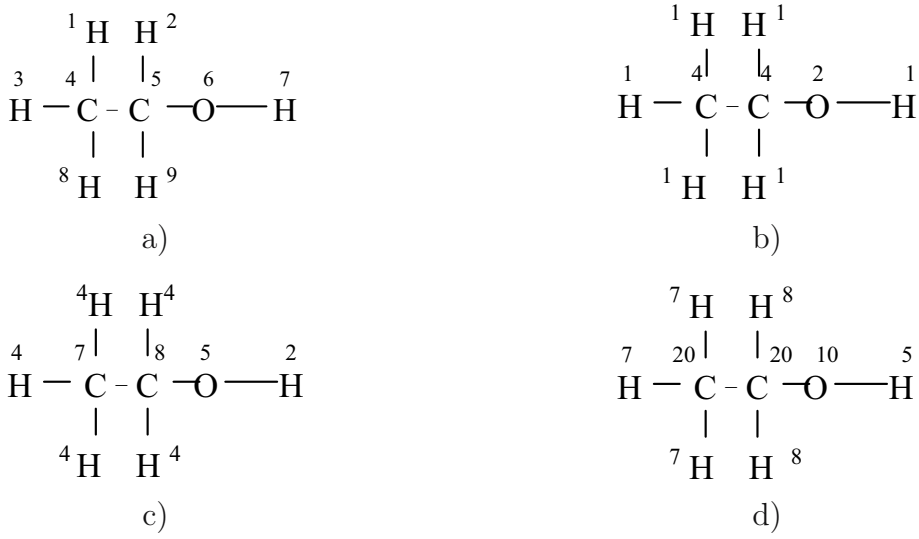

Fig. 2 Local invariants of the LHFG of ethanol: (a) arbitrary numbering of the vertices; (b) ^{0}EC; (c) ^{1}EC; (d) ^{2}EC.

The starting value of each correlation weight was 1. The Monte Carlo iterative optimization procedure [29, 46, 47] was used to derive the best values of correlation weights (i.e. CW(a_k) and CW(^{x}EC). The "best values of CW's" were the largest possible correlation coefficient between the $\log(10^6/C_{50})$ values of the molecular set and the molecular descriptor DCW. The molecular descriptor was defined on the basis of the optimized correlation weights and it was then used to derive all the relations with $\log(10^6/C_{50})$ values,

employing the least-squares method of the linear regression

$$\log(10^6/C_{50}) = C_0 + C_1 \cdot DCW(a,^x EC) \tag{2}$$

Several partitions of the data set into a training set and a test set were done and the corresponding statistical parameters in each case calculated. We found no significant differences among them. Then, we chose one of these partitions and the selected set is listed in Table 1. The optimization of correlation weights was done using a program developed by one of the authors (AAT). Least-squares linear regression analyses were done employing a standard computer program, Sigma Plot, Excel and Origin professional software. The statistical qualities of equations were judged by parameters r (correlation coefficient), F (variance ratio), s (standard error of estimate) and AVRES (average of absolute values of residuals).

3 Results and discussion

In a recent study, a rather simple and direct application of previous formulation was applied to calculate biological activities [48]. However, the method can be employed in a more comprehensive way in order to extent its capabilities .

Table 2 lists results of OCWLGI of three probes based on optimization of correlation weights of different chemical elements present, CW(a_k), together with the CW(^{x}EC). The correlation weights of first OCWLGI probe are presented in Table 3. The results of the optimized correlation weights vary for each trial due to the statistical nature of the Monte Carlo method. However, differences are not significant for different probes corresponding to a given descriptor, as shown in Table 2, for the three cases reported for each OCWLGI based on the DCW(a,^{x}EC, x = 0, 1, 2).

Table 2 shows that best QSAR model for the anti-HIV-1 activity takes place in case of the [DCW(a,^{1}EC$_k$). Table 3 lists relevant CW values for calculating the descriptor.

Amongst the molecular descriptors DCW(a,^{x}EC), with x = 0, 1, 2, DCW(a,^{1}EC) gives the best correlation between the descriptor and the [log(10^6/C$_{50}$)]. The analysis of data displayed in Table 3 shows an agreement between experimental and theoretical data. The AVRES for training and test sets, are similar, although, as expected, it is slightly smaller for the training set. Since results for the test set are true predictions, this similarity is encouraging. Among the 20 molecules comprised in the test set there were only two of them that had a relatively large deviation (molecules No. 14 and 11).

On comparison of theoretical results from our calculations with those previously published for an identical molecular set [22], we found that our results present a stronger and clearer prediction. In fact, although statistical parameters associated with previous regression equations (see equations 8-11 in Ref. [22]) and ours are nearly the same, numerical data corresponding to our test set are true predictions, while those reported by Politzer *et al.* correspond to two separate sets (TIBO and HEPT derivatives) both being considered as training sets. Besides, Politzer *et al*'s equations employ three and four independent variables/molecular descriptors for TIBO and HEPT derivatives, respectively,

while our regression equation depends upon just one variable (i.e. DCW(a, ^{x}EC)

4 Conclusions

Utilizing the Morgan extended connectivity index of first order correlation weights of local graph invariants makes up a reasonable good approximation to predict the anti-HIV-activities for two representative molecular sets namely, TIBO and HEPT classes. The power of flexible indices based on the optimization of correlation weights of local graph invariants over the approach based on computed molecular surface electrostatic potentials [22] is shown. The advantage of the present model is grounded in one-variable regression equation while the latter method resorts to the use of several variables to attain comparable results. Moreover, the present scheme does not require any rather complex calculation of diverse descriptors and statistical analysis for proper selection of descriptors and inter-correlation among them. The model merits additional assessment on exploring quantitative structure property (activity) of different physicochemical properties and biological data using several different local invariants to justify its suitability in modeling studies. Work is in progress for performing such calculations and results will be published elsewhere in the forthcoming future.

References

[1] A.F. Jalbout and X. Li: "Anti-HIV-1 inhibitors of various molecules using principles of connectivity", *J. Mol. Struct. (Theochem)*, Vol. 663, (2003), pp. 19–23.

[2] H. Yuan and A.L. Parrill: "QSAR Develpment to describe HIV-1 integrase inhibition", *J. Mol. Struct. (Theochem),* Vol. 529, (2000), pp. 273–282.

[3] V.P. Solov'ev and A. Varnek: "Anti-HIV activity of HEPT, TIBO, and cyclic urea derivatives: Structure-property studies, focused combinatorial library generation, and hits selection using substructural molecular fragments method", *J. Chem. Inf. Comput. Sci.,* Vol. 43, (2003), pp. 1703–1719.

[4] M.T. Makhija and V.M. Kulkarni: "Eigen value Analysis of HIV-1 integrase inhibitors", *J. Chem. Inf. Comput. Sci.,* Vol. 41, (2001), pp. 1569–1577.

[5] *SciFinder Scholar, version 2000.1*, American Chemical Society, Washington, DC, 2000, http://www.cas.org/SCIFINDER/SCHOLAR/index.html.

[6] R. Garg, S.P. Gupta, H. Gao, M.S. Babu, A.K. Debnath and C. Hansch: "Comparative Quantitative Structure-Activity Relationship studies on anti-HIV drugs", *Chem. Rev.,* Vol. 99, (1999), pp. 3525–3602.

[7] M. Jalali-Heravi and F. Parastar: "Use of artificial neural networks in a QSAR study of anti-HIV activity for a large group of HEPT derivatives", *J. Chem. Inf. Comput. Sci.,* Vol. 40, (2000), pp. 147–154.

[8] M.H. Knaggs, C. McGuigan, S.A. Harris, P. Heshmati, D. Cahard, I.H. Gilbert and J. Balzarini: "A QSAR study investigating the effect of L-Alanine ester variation

on the anti-HIV activity of some phosphoramidate derivatives of d4T", *Bioorg. Med. Chem. Lett.*, Vol. 10, (2000), pp. 2075–2078.

[9] J.M.J. Tronchet, M. Grigorov, N. Dolatshahi, F. Moriaud and J.A. Weber: A QSAR study confirming the heterogeneity of the HEPT derivative series regarding their interaction with HIV reverse transcriptase", *Eur. J. Med. Chem.,* Vol. 32, (1997), pp. 279–299.

[10] J. Houskonen: "QSAR Modeling with the electrotopological state: TIBO derivatives", *J. Chem. Inf. Comput. Sci.,* Vol. 41, (2001), pp. 425–429.

[11] H.H. Maw and L.H. Hall: "E-State Modeling of HIV-1 protease inhibitor binding independent of 3D information", *J. Chem. Inf. Comput. Sci.,* Vol. 42, (2002), pp. 190–298.

[12] E. Gancia, G. Bravi, P. Mascagni and A. Zaliani: "Global 3D-QSAR Methods: MS-WHIM and autocorrelation", *J. Comput. Aid. Mol. Des.,* Vol. 14, (2000) pp. 293–306.

[13] C.T. Klein, L. Lawtrakul, S. Hannongbua and P. Wolschann: "Accessible charges in Structure- Activity Relationships. A study on HEPT-Based HIV-1 RT inhibitors", *Sci. Pharm.,* Vol. 68, (2000), pp. 25–40.

[14] O.A. Santos-Filho and A.J. Hopfinger: "The 4D-QSAR paradigm: Application to a novel set of nonpetidic HIV protease inhibitors", *Quant. Struct.-Act. Rel.,* Vol. 21, (2002), pp. 369–381.

[15] D.B. Kireev, J.R. Chretien and O.A. Raevsky: "Molecular Modeling and Quantitative Structure- Activity studies of anti-HIV-1,2-heteroarylquinoline-4-amines", *Eur. J. Med. Chem.,* Vol. 30, (1995), pp. 395–402.

[16] S. Hennongbua, K. Nivesanond, L. Lawtrakul, P. Pungpo and P. Wolschann: "3D-Quantitative Structure-Activity Relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on *ab initio* calculations", *Chem. Inf. Comput. Sci.,* Vol. 41, (2001), pp. 848–855.

[17] P.R.N. Jayatilleke, A.C. Nair, R. Zauhar and W.J. Welsh: "Computational studies on HIV-1 protease inhibitors: Influence of calculated inhibitor-enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models", *J. Med. Chem.,* Vol. 43, (2000), pp. 4446–4451.

[18] A.K. Debnath: "Three-dimensional Quantitative Structure-Activity Relationships study on cyclic urea derivatives as HIV-1 protease inhibitors: Application to comparative molecular field analysis", *J. Med. Chem.,* Vol. 42, (1999), pp. 249–259.

[19] J.K. Buolamwini and H. Assefa: "CoMFA and CoMSIA 3D QSAR and docking studies on conformationally-restrained cinnamoyl HIV-1 integrase inhibitors: Exploration of a binding mode at the active site", *J. Med. Chem.,* Vol. 45, (2002), pp. 841–852.

[20] T. Mickle and V. Nair: "Anti-human immunodeficiency virus activities of nucleosides and nucleotides: Correlation with molecular electrostatic potential data", *Antimicrob. Agents Ch.*, Vol. 44, (2000), pp. 2939–2947.

[21] T. Mickel and V. Nair: "Predictive QSAR analysis of anti-HIV agents", *Drugs Future,* Vol. 25, (2000), pp. 393–400.

[22] O.G. Gonzáles, J.S. Murray, Z. Peralta-Inga and P. Politzer: "Computed molecular surface electrostatic potentials of two groups of riverse transcriptase inhibitors: Relationships to anti- HIV-1 activities", *Int. J. Quantum Chem.*, Vol. 83, (2001), pp. 115–124.

[23] M. Firpo, L. Gavernet, E.A. Castro and A.A. Toropov: "Maximum topological distances based indices as molecular descriptors for QSPR. Part 1. Application to alkyl benzenes boiling points", *J. Mol. Struct. (Theochem),* Vols. 501-502, (2000), pp. 419–425.

[24] E.A. Castro, M. Tueros and A.A. Toropov: "Maximum topological distances based indices as molecular descriptor for QSPR. 2 – Application to aromatic hydrocarbons, *Computers and Chemistry,* Vol. 24, (2000) pp. 571–576.

[25] A. Mercader, E.A. Castro and A.A. Toropov: "Maximum topological distances based indices as molecular descriptors for QSPR. 4 – Modeling the enthalpy of formation of hydrocarbons from elements", *Int. J. Mol. Sci.*, Vol. 2, (2001), pp. 121–134.

[26] A. Toropov, A. Toropova, T. Ismailov and D. Bonchev: "3D weighting of molecular descriptors for QSPR/QSAR by the Method of Ideal Symmetry (MIS). 1. Application to boiling points of alkanes", *J. Mol. Struct. (Theochem),* Vol. 424, (1998), pp. 237–247.

[27] G. Krenkel, E.A. Castro and A.A. Toropov: "3D and 4D Molecular models derived from the ideal symmetry method: Prediction of alkanes normal boiling points", *Chem. Phys. Lett.,* Vol. 355, (2002), pp. 517–528.

[28] A.A. Toropov and A.P. Toropova: "Method of ideal symmetry in four-dimensional space: Implementation in the QSPR studies on the thermochemistry of complex compounds", *Russ. J. Coord. Chem.,* Vol. 23, (1997), pp. 741–747.

[29] A.A. Toropov and A.P. Toropova: "Optimization of correlation weights of the local graph invariants: Use of the enthalpies of formation of complex compounds for the QSPR modeling", *Russ. J. Coord. Chem.,* Vol. 24, (1998), pp. 81–85.

[30] A.A. Toropov and A.P. Toropova: "QSPR Modeling of stability of complexes of adenosine phosphate derivatives with metals absent from the complexes of the teaching access", *Russ. J. Coord. Chem.*, Vol. 27, (2001), pp. 574–578.

[31] A.A. Toropov and A.P. Toropova: "Modeling of acyclic compounds normal boiling points by correlation weighting of nearest neighboring codes", *J. Mol. Struct. (Theochem)*, Vol. 581, (2002), pp. 11–15.

[32] A.A. Toropov and A.P. Toropova: "Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graph of local invariants", *J. Mol. Struct. (Theochem),* Vol. 538, (2001), pp. 287–293.

[33] A.A. Toropov and A.P. Toropova: "QSAR Modeling of mutagenicity based on graph of atomic orbitals", *Internet Electron. J. Mol. Des.*, Vol. 1, (2002), pp. 109–113, http:// www. biochempress.com

[34] A.A. Toropov and A.P. Toropova: "QSAR Modeling of toxicity on optimization of correlation weights of Morgan extended connectivity", *J. Mol. Struct. (Theochem),* Vol. 578, (2002) pp. 129–134.

[35] A.P. Toropova, A.A. Toropov, M.M. Ishankhodzhaeva and N.A. Parpiev: "QSPR Modeling of stability constants of coordination compounds by optimization of correlation weights of local Ggraph invariants", *Russ. J. Inorg. Chem.,* Vol. 45, (2000) pp. 1057–1059.

[36] G. Krenkel, E.A. Castro and A.A. Toropov: "Improved molecular descriptors to calculate boiling points based on the optimization of correlation weights of local graph invariants", *J. Mol. Struct. (Theochem)*, Vol. 542, (2001), pp. 107–113.

[37] G. Krenkel, E.A. Castro and A.A. Toropov: "Improved molecular descriptors based on the optimization of correlation weights of local graph invariants", *Int. J. Mol. Sci.*, Vol. 2, (2001), pp. 57–65.

[38] A. Mercader, E.A. Castro and A.A. Toropov: "QSPR Modeling of the enthalpy of formation from elements by means of correlation weighting of local invariants of atomic orbital molecular graphs", *Chem. Phys. Lett.,* Vol. 330, (2000), pp. 612–623.

[39] P.R. Duchowicz, E.A. Castro and A.A. Toropov: "Improved QSPR analysis of standard entropy of acyclic and aromatic compounds using optimized correlation weights of linear graph invariants", *Computers & Chemistry,* Vol. 26, (2002), pp. 327–332.

[40] P.J. Peruzzo, D.J.G. Marino, E.A. Castro and A.A. Toropov: "Calculation of pK values of flavylium salts from the optimization of correlation weights of local graph invariants", *J. Mol. Struct. (Theochem),* Vol. 572, (2001), pp. 53–60.

[41] M. Randic: "On computation of optimal parameters for multivariate analysis of structure- property relationship", *J. Comput. Chem.,* Vol. 12, (1991), pp. 970–980.

[42] M. Randic: "Novel graph theoretic approach to heteroatoms in Quantitative Structure-Activity Relationships", *Chemom. Intell. Lab. Syst.,* Vol. 10, (1991), pp. 213–227.

[43] M. Randic: "Resolution of ambiguities in structure-property studies by use of orthogonal descriptors", *J. Chem. Inf. Comput. Sci.,* Vol. 31, (1991), pp. 311–320.

[44] W. Ho, M.J. Kukla, H.J. Breslin, D.W. Ludovici, P.P. Grous, C.J. Diamond, M. Miranda, J.D. Rodgers, C.Y. Ho, E. De Clercq, R. Pauwels, K. Andries, M.A.C. Janssen and P.A.J. Janssen: "Synthesis and Anti-HIV-1 Activity of 4,5,6,7- Tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TlBO) Derivatives", *J. Med. Chem.*, Vol. 38, (1995), p. 794 and earlier references cited.

[45] M. Baba, S. Shigeta, H. Tanaka, T. Miyasaka, M. Ubasawa, K. Umezu, R.T. Walker, R. Pauwles and E. De Clerq: *Antiviral Res.,* Vol. 17, (1992), p. 245.

[46] A.A. Toropov, A.P. Toropova, N.L. Voropaeva, I.N. Ruban and S.Sh. Rashidova: "Generalized zero-order molecular connectivity index: Enthalpies of crystalline aquo and ammino complexes in QSPR modeling", *Russ. J. Coord. Chem.,* Vol. 24, (1988), pp. 525–529.

[47] A.A. Toropov, N.L. Voropaeva, I.N. Ruban and S.Sh. Rashidova: "Quantitative Structure- Property Relationships for binary polymer solvent systems: Correlation weighting of the local invariants of molecular graphs", *Polym. Sci. Ser. A,* Vol. 41, (1999), pp. 975–985.

[48] E.A. Castro, F. Torrens, A.A. Toropov, I.V. Nesterov and O.M. Nabiev: "QSAR

Modeling of anti-HIV-1 activities by optimization of correlation weights of local graph invariants", *Mol. Simulat.,* Vol. 30, (2004), pp. 691–696.

Training set

No	X	Y	Z	R	Type	Exp.	Calc.	Exp.-Calc.
1	8-Br	5-CH_3	S	DMA	TIBO	8.52	8.28	0.25
2	8-F	5-CH_3	S	DMA	TIBO	8.24	6.94	1.30
3	8-Cl	7-CH_3	S	DMA	TIBO	7.92	7.76	0.16
4	8-CH_3	5-CH_3	S	DMA	TIBO	7.87	7.63	0.24
5	9-Cl	5-CH_3	S	DMA	TIBO	7.47	7.76	-0.29
6	8-Cl	H	S	DMA	TIBO	7.34	7.42	-0.08
7	8-I	5-CH_3	S	DMA	TIBO	7.32	7.55	-0.23
8	8-CN	5-CH_3	S	DMA	TIBO	7.25	6.84	0.41
9	8-I	5-CH_3	O	DMA	TIBO	7.06	7.25	-0.19
10	H	5-CH_3	O	DMA	TIBO	7.01	6.11	0.90
11	9-Cl	7-CH_3	O	DMA	TIBO	6.80	6.58	0.22
12	9-CF_3	5-CH_3	S	DMA	TIBO	6.31	6.56	-0.25
13	8-CH_3	5-CH_3	O	DMA	TIBO	6.00	6.45	-0.45
14	8-CN	5-CH_3	O	DMA	TIBO	5.94	6.39	-0.45
15	9-NO_2	5-CH_3	S	CPM	TIBO	5.61	5.60	0.01
16	10-OCH_3	5-CH_3	S	DMA	TIBO	5.33	5.39	-0.06
17	9-CF_3	5-CH_3	O	DMA	TIBO	5.23	5.38	-0.15
18	H	7-CH_3	O	DMA	TIBO	4.92	6.11	-1.19
19	H	5-CH_3	O	$CH_2C(C_2H_5)=CH_2$	TIBO	4.43	4.43	0.01
20	H	5-CH_3	O	$CH_2CH_2CH=CH_2$	TIBO	4.30	4.45	-0.15
21	H	5-CH_3	O	C_3H_7	TIBO	4.22	3.62	0.60
22	H	5-CH_3	O	$CH_2CH=CH_2$	TIBO	4.15	4.09	0.06
23	H	4-$CH(CH_3)_2$	O	C_3H_7	TIBO	4.13	4.79	-0.66
24	8-NH_2	5-CH_3	O	CPM	TIBO	3.07	3.05	0.03
25	O	C_2H_5	C_6H_5	3,5-$(CH_3)_2$	HEPT	8.62	8.83	-0.21
26	O	$CH(CH_3)_2$	CH_2OH	3,5-$(CH_3)_2$	HEPT	8.48	7.90	0.58
27	O	C_2H_5	C_6H_5	H	HEPT	8.31	8.16	0.16
28	S	C_2H_5	CH_3	3,5-$(CH_3)_2$	HEPT	8.25	8.56	-0.31
29	O	C_2H_5	CH_3	3,5-$(CH_3)_2$	HEPT	8.21	7.38	0.84
30	O	$CH(CH_3)_2$	CH_3	H	HEPT	8.09	7.47	0.63
31	S	$CH(CH_3)_2$	CH_3	H	HEPT	7.92	8.65	-0.73
32	O	C_2H_5	CH_3	H	HEPT	7.66	6.70	0.96
33	S	C_2H_5	CH_3	H	HEPT	7.59	7.88	-0.29
34	O	C_2H_5	CH_2OH	H	HEPT	6.92	6.45	0.47
35	O	CH_3	CH_2OH	3,5-$(CH_3)_2$	HEPT	6.59	6.73	-0.14
36	O	CH_3	CH_2OH	3-CH_3	HEPT	5.59	6.39	-0.80
37	O	CH_3	C_2H_5	H	HEPT	5.52	6.70	-1.18

AVRES = 0.42

Table 1 Observed [22] and calculated (Eq. 2 with x = 1) $\log(10^6/C_{50})$ values of TIBO and HEPT structural types of reverse transcriptase inhibitors.

Test set

No	X	Y	Z	R	Type	Exper.	Calc.	Exp.-Calc.
1	8-Cl	5-CH_3	S	DMA	TIBO	8.37	7.76	0.61
2	9-F	5-CH_3	S	DMA	TIBO	7.60	6.94	0.66
3	H	5,7-CH_3	S	DMA	TIBO	7.38	7.63	-0.25
4	H	5-CH_3	S	DMA	TIBO	7.36	7.29	0.07
5	8-Br	5-CH_3	O	DMA	TIBO	7.33	7.10	0.24
6	H	7-CH_3	S	DMA	TIBO	7.11	7.29	-0.18
7	8-Cl	7-CH_3	O	DMA	TIBO	6.84	6.58	0.26
8	9-Cl	H	S	DMA	TIBO	6.80	7.42	-0.62
9	H	7-CH_3	S	C_3H_7	TIBO	5.61	4.80	0.81
10	H	5-CH_3	O	DMA	TIBO	5.48	6.11	-0.63
11	10-OCH_3	5-CH_3	O	DMA	TIBO	5.18	4.21	0.97
12	9-NO_2	5-CH_3	O	CPM	TIBO	4.48	4.42	0.06
13	H	5-CH_3	O	$CH_2(CH_2)_2CH_3$	TIBO	4.00	4.39	-0.39
14	H	5-CH_3	O	$CH_2CO(O)CH_3$	TIBO	3.07	2.03	1.04
15	O	$CH(CH_3)_2$	C_6H_5	H	HEPT	8.47	8.92	-0.45
16	O	C_2H_5	CH_2OH	3,5-$(CH_3)_2$	HEPT	7.80	7.13	0.67
17	O	$CH(CH_3)_2$	CH_2OH	H	HEPT	7.14	7.22	-0.08
18	O	CH_3	C_6H_5	H	HEPT	7.03	7.75	-0.72
19	O	CH_3	CH_3	H	HEPT	6.48	6.30	0.18
20	O	CH_3	CH_2OH	H	HEPT	5.19	6.05	-0.86

AVRE = 0.49

Table 1 (continued) Observed [22] and calculated (Eq. 2 with x = 1) $\log(10^6/C_{50})$ values of TIBO and HEPT structural types of reverse transcriptase inhibitors.

Probe	Training Set n = 37 R	s	F	Test Set n = 20 R	s	F	All compounds n = 57 R	s	F
	OCWLGI based on the DCW(a,^{0}EC)								
1	0.8712	0.754	110	0.9121	0.625	89	0.8834	0.706	195
2	0.8660	0.768	105	0.9075	0.639	84	0.8784	0.720	186
3	0.8664	0.767	105	0.9012	0.649	78	0.8773	0.722	184
	OCWLGI based on the DCW(a,^{1}EC)								
1	0.9320	0.557	231	0.9343	0.590	124	0.9295	0.564	349
2	0.9317	0.558	230	0.9357	0.576	127	0.9303	0.559	354
3	0.9321	0.557	232	0.9359	0.588	127	0.9300	0.562	352
	OCWLGI based on the DCW(a,^{2}EC)								
1	0.9499	0.480	324	0.8990	0.730	76	0.9281	0.574	342
2	0.9498	0.481	322	0.9067	0.705	83	0.9308	0.563	356
3	0.9499	0.480	323	0.9010	0.705	78	0.9300	0.563	352

Table 2 Results of OCWLGI based on the DCW(a,^{0}EC), DCW(a,^{1}EC), and DCW(a,^{2}EC).

LHFG Local invariants Chemical elements	Correlation weights a_k, CW(a_k)
C	-0.575
N	-3.178
H	0.150
Br	1.298
S	2.333
F	-0.266
Cl	0.695
I	1.482
O	0.953
Morgan extended connectivity of first order ^{1}EC, CW(^{1}EC)	
0002	3.390
0003	1.359
0004	0.275
0005	-3.775
0006	0.071
0007	-0.509
0008	1.408
0009	0.775
0010	0.625
0011	2.066
0012	1.906
0013	2.933

Table 3 Correlation weights of first OCWLGI based on the DCW(a,^{1}EC).