In Silico Studies toward the Discovery of New Anti-HIV Nucleoside Compounds with the Use of TOPS-MODE and 2D/3D Connectivity Indices. 1. Pyrimidyl Derivatives

Ernesto Estrada,* Santiago Vilar, Eugenio Uriarte, and Yaquelin Gutierrez

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela, Santiago de Compostela 15706, Spain

Received April 26, 2002

Computational approaches are developed to design or rationally select, from structural databases, pyrimidyl nucleosides with anti-HIV activity. A data set of 141 nucleoside derivatives was selected from literature, and a discriminant function was derived with the use of TOPS-MODE descriptors. The model is able to classify correctly 83% of the compounds in a training set and 88.5% in a cross-validation set. The use of an external prediction set selected from the most recent literature proved that the model has good predictive ability, with a good classification of 85% of the compounds in this set. This model permitted the structural interpretation of the anti-HIV activity of these nucleoside analogues. This interpretation is formulated as several rules concerning the influence of several structural features on the activity/inactivity of such compounds. A QSAR model for the most active compounds was developed with the combined use of 2D and 3D connectivity indices. This model explains 88% of the variance in the activity of these compounds in MT4 assay. The combination of both models will permit the selection of pyrimidyl nucleoside leads and their optimization to improve the potency of the selected ones.

INTRODUCTION

Nucleoside derivatives are well-established today as effective antiviral agents in spite of the toxicity generally associated with them.[1−5] In fact, the first drugs approved for clinical use in the treatment of HIV infections were 2',3'-dideoxynucleosides, i.e., AZT, zidovudine; DDC, zalcitabine; and DDI, didanosine.[1] These compounds are inhibitors of the HIV reverse transcriptase (HIV RT), which has been one of the most studied viral targets in the development of anti-HIV drugs. The chemical modification of the nucleoside moiety permits one to convert a normal substrate for nucleic acid synthesis into a potent inhibitor of the HIV replication. Thus, a great diversity of nucleoside derivatives has been explored in search of HIV inhibitors.[3] They include modifications on the base, on the sugar, and the separation between them.[3]

In this great pool of nucleoside derivatives there are some promising drug candidates for the treatment of HIV infections besides several drugs used today in clinics. However, several structural modifications of the nucleoside moiety have not produced active compounds against HIV or the activities found are only moderate.[3] For this reason it is necessary to develop computational methods permitting theoretical evaluations of anti-HIV activity for all possible structural modifications feasible over the nucleoside structure before these compounds are synthesized in the lab.[6−9] This "in silico" world of data, analysis, hypothesis, and models that reside inside a computer is alternative to the "real" world of synthesis and screening of nucleosides in the laboratory. The distinguishable characteristic of these theoretical models focused on virtual screening of nucleosides is that they have to be based on a large and structurally variable data set of compounds. This will guarantee that the model "learn" most of the possible structural changes over the nucleoside moiety for a more general applicability of the model. By this mean, the molecular descriptors to be used have to be applicable to large and variable data sets in a fast and efficient way. 2D (topological) molecular descriptors[10] appear as good candidates in this direction because they describe the molecular structure in an efficient way and are very easy to compute for large databases of structures. Their main disadvantage is that they do not consider explicitly the 3D structural features of molecules, which are of relevant importance for the description of biological activity in many cases. However, in the search of novel nucleoside derivatives with anti-HIV activity great topological changes are expected in the main framework of the molecule. For instance, the changes in the size of the sugar moiety, hexacyclic, pentacyclic, tetracyclic, tricyclic, or the change from cyclic to acyclic, or the change from heterocyclic to carbocyclic, are examples of drastic topological changes on the nucleoside structure. On the other hand, the identification of the different chiral centers in the nucleoside framework permits the use of indicator (dummy) variables for distinguishing among the different stereoisomers in the molecule. These two factors together with the easy computation and applicability to large and variable data sets make the topological substructural molecular design (TOPS-MODE) approach as well as 2D and 3D connectivity indices excellent choices for the current study.

METHODS AND MATERIALS

TOPS-MODE Approach. The *TOP*ological *S*ub-Structural *MO*lecular *DE*sign (TOPS-MODE) approach will be used here for the virtual screening of anti-HIV nucleoside

* Corresponding author fax: 34-981-594912; e-mail: estrada66@yahoo.com.

10.1021/ci0255331 CCC: $22.00
Published on Web 07/27/2002

Table 1. Standard and Dipole Moments Used in the Calculation of the TOPS-MODE Descriptors

bond	dipole (D)[a]	bond	dipole (D)[a]	bond	dipole (D)[a]
C−C	0	C−N	0.40	C−S	2.95
C−F	1.51	C=N	0.90	C=S	2.80
C−Cl	1.56	C∞N	3.60	N−O	0.30
C−Br	1.48	C−O	0.86	N=O	2.00
C−I	1.29	C=O	2.40		

[a] Bond dipole taken from ref. [59].

compounds. The general principles of this approach have been explained with some detail elsewhere.[11−15] Some of the applications of this approach reported in the literature in drug design and QSAR studies covered the areas of antifungal activity of phenols against *Aspergillus niger*,[13] sedative/hypnotic activity of diverse data set of compounds,[16] anticancer activity of diverse data set of compounds,[17] anticonvulsant activity of a diverse data set of compounds,[18] and toxicity of nitrobenzenes against *Tetrahymena pyriformis*[19] as well as the neurotoxicity of solvent in rats and mice.[20]

Data Sets. A data set of 141 pyrimidyl nucleoside derivatives was collected from the literature for the present study.[21−58] This data set is divided into two subgroups. A set of 65 compounds (21 active and 44 inactive) was used as the training set for developing the classification model with TOPS-MODE. The second subgroup of 56 compounds (12 active and 44 inactive) was used as a cross-validation set for testing the predictive capacity of the discrimination model. Another data set of pyrimidyl nucleoside derivatives was used as an external prediction set. It is formed by 20 compounds (3 active and 17 inactive) selected from recent literature and was never used in the training or cross-validation process.[21,24,29,36,41,46−52,55−58]

The criterion of activity selected in this study is based on the effective concentration EC_{50} on MT4 assay. As we are interested in selecting only very active compounds against the selection of moderate or poorly active ones, we have taken as cutoff value of activity an EC_{50} of 10 μM. That is, active compounds are those having $EC_{50} \leq 10$ μM and inactive ones are those having $EC_{50} > 10$ μM.

Structural Calculations. Standard bond dipole moments are used in the current work for weighting the different bonds of nucleoside derivatives. The values used in this study are given in Table 1.[59]

Three indicator variables I_1, I_2, and I_3 were used for distinguishing chiral centers at positions 1′, 3′, and 4′ in the sugar ring, respectively. These variables take values of 1, −1, or 0 if the center has configuration R, S, or not chirality, respectively.

Discriminant Analysis. The use of the statistical analysis will permit one to find a classifier function by using any discrimination statistical technique, such as linear discriminant analysis

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4 + a_5\mu_5 + a_6\mu_6 + b \quad (1)$$

where P is the biological property, μ_k is the kth spectral moment, i.e., the sum of the diagonal entries of the kth power of the bond matrix, and the a_k's are the coefficients obtained by the linear discriminant analysis (LDA).

LDA is used in order to generate the classifier function on the basis of the simplicity of the method. All statistical analysis were carried out with the STATISTICA package. The discriminant function is obtained by using the stepwise discriminant analysis with Fisher ratio F to enter and to remove 1 and 0, respectively. The tolerance parameter used was the default value for the minimum acceptable tolerance, which is 0.01. The linear regression analysis uses similar parameters to generate the QSAR model.

To test the quality of the discriminant functions derived we used the Wilks' λ and the Mahalanobis distance. The classification of cases was carried out by means of the posterior classification probabilities, which is the probability that the respective case belongs to a particular group, i.e., active or inactive.

In developing this classification function the values of 1 and −1 were assigned to active and inactive compounds. By using this model one compound is classified as active if $C \geq 0.2$ and as inactive if $C \leq -0.2$. The interval between −0.2 to 0.2 is left for not classified compounds. However, we will use the a posteriori probabilities instead of these cutoff values in order to classify the compounds as active/inactive.

Computation of Fragment Contributions. The general algorithm followed is as follows. First, select the substructure whose contribution to the moments will be determine. Then generate all fragments (subgraphs) which are contained in the corresponding substructure and calculate the spectral moments for both the substructure and all their fragments. The contribution of the substructure to the spectral moments is finally obtained as the difference between the spectral moments of the substructure and all those from their fragments. Substitute these contributions into the quantitative model developed to describe the property studied, e.g., model (1), obtaining the quantitative contribution of the different fragments to P. These calculations are made automatically by the computer software TOPS-MODE.

2D/3D Connectivity Indices QSAR Model. 2D (topological)[60−66] and 3D (topographic)[67−70] connectivity indices were calculated for describing quantitatively the anti-HIV activity of some nucleosides via QSAR models. Topographic 3D connectivity indices are able to distinguish among different molecular configurations and conformations.[67−70] Details on the way these descriptors are defined and calculated are provided in extension elsewhere.[67−70] Consequently we have used full-geometry optimization of the molecular structures with the semiempirical quantum chemical method AM1[71] as implemented in MOPAC 6.0 in order to calculate these descriptors.[72] A conformational sampling for all the structures studied here is a very time-consuming procedure that was skipped here by starting the calculations from the preferred conformations according to previous quantum chemical studies.[5] The computer system MODEST[73] was used to calculate these descriptors, whose definitions and symbols are given in Table 2. The QSAR models were selected from the best stepwise linear regression models developed with STATISTICA.[74] In testing these models the leave-one-out cross-validation error was used, and the presence of statistical outliers was checked by the following statistical tests: residuals, standardized residuals, and studentized residuals. One compound was considered as outlier if its residual lies three standard deviations from

Table 2. Symbols for Topological, Topographic and Quantum Chemical Descriptors and Their Definitions

$^h\chi_p$	path connectivity index of order $h = 0-6$
$^h\chi_C$	cluster connectivity index of order $h = 3-6$
$^h\chi_{pC}$	path-cluster connectivity index of order $h = 4-6$
$^h\chi_p^v$	valence path connectivity index of order $h = 0-6$
$^h\chi_C^v$	valence cluster connectivity index of order $h = 3-6$
$^h\chi_{pC}^v$	valence path-cluster connectivity index of order $h = 4-6$
$^h\epsilon_p$	path bond connectivity index of order $h = 1-6$
$^h\epsilon_C$	cluster bond connectivity index of order $h = 3-6$
$^h\epsilon_{pC}$	path-cluster bond connectivity index of order $h = 4-6$
$^h\Omega_p$	path bond-order-based topographic connectivity index of order $h = 0-6$
$^h\Omega_C$	cluster bond-order-based topographic connectivity index of order $h = 3-6$
$^h\Omega_{pC}$	path-cluster bond-order-based topographic connectivity index of order $h = 4-6$
$^h\Omega_p(q)$	path charge-based topographic connectivity index of order $h = 0-6$
$^h\Omega_C(q)$	cluster charge-based topographic connectivity index of order $h = 3-6$
$^h\Omega_{pC}(q)$	path-cluster charge-based topographic connectivity index of order $h = 4-6$
$^h\Omega_p^c(q)$	path hydrogen-corrected charge-based topographic connectivity index of order $h = 0-6$
$^h\Omega_C^c(q)$	cluster hydrogen-corrected charge-based topographic connectivity index of order $h = 3-6$
$^h\Omega_{pC}^c(q)$	path-cluster hydrogen-corrected charge-based topographic connectivity index of order $h = 4-6$
$^h\epsilon_p(\rho)$	path bond-order-based topographic bond connectivity index of order $h = 1-6$
$^h\epsilon_C(\rho)$	cluster bond-order-based topographic bond connectivity index of order $h = 3-6$
$^h\epsilon_{pC}(\rho)$	path-cluster bond-order-based topographic bond connectivity index of order $h = 4-6$

the mean of the residuals and/or the standardized and studentized residuals exceed 3 in absolute value.

The interrelation among the molecular descriptors makes the interpretation of the QSAR model difficult. To overcome this difficulty an approach based on the orthogonalization of the descriptors has been introduced in the literature. The orthogonalization process of molecular descriptors was introduced by Randić 10 years ago as a way to improve the statistical interpretation of the models built by using interrelated indices.[75−79] The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of its collinearity with other variables previously included in the model. It is known that the interrelatedness among the different descriptors can result in highly unstable regression coefficients, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficients in a model. However, in some cases strongly interrelated descriptors can enhance the quality of a model because the small fraction of a descriptor which is not reproduced by its strongly interrelated pair can provide positive contributions to the modeling.

The Randić method of orthogonalization has been described in details in several publications.[75−79] Thus, we will give only a general overview here. The first step in orthogonalizing the molecular descriptors is to select the appropriate order of orthogonalization, which in this case is the order in which the variables were selected in the forward stepwise search procedure of the linear regression analysis. The first variable ($v1$) is taken as the first orthogonal descriptors $^1O(v1)$ and the second one is orthogonalized with respect to it by taking the residual of its correlation with $^1O(v1)$. The process is repeated until all variables are completely orthogonalized, and the orthogonal variables are then used to obtain the new model.

RESULTS AND DISCUSSION

Classification Model. The linear classification model obtained is given below together with the statistical parameters

$$Class = -3.8605\mu_1 - 0.6091\mu_2 + 0.5871\mu_3 - 7.027\cdot10^5\mu_9 + 2.9071\cdot10^8\mu_1\mu_{12} + 0.1682I_1 - 2.5934I_2 - 0.8707I_3 + 11.766 \quad (2)$$

$$N = 65,\ \lambda = 0.64,\ D^2 = 2.58$$

where λ is the Wilks' statistics and D^2 is the squared Mahalanobis distance.

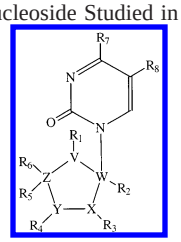
Model (2) classifies correctly 85.7% of active and 81.8% of inactive compounds in the training set, for a global good classification of 83.1%. The percentages of false actives and false inactives in the training set are 6.2% (4/65) and 4.6% (3/65), respectively. False actives are those inactive compounds that the model classifies as actives, and the false inactive are active compounds classified as inactives by the model. In Tables 3 and 4 we give the classification of compounds in the training set together with their posterior probabilities calculated from the Mahalanobis distance. Four compounds were not classified by this model because their percentage of active and inactive differ by less than 5%. This represents 6.2% of unclassified (U) compounds, whose not inclusion in the general statistics increases the percentage of good classification of the model to 88.5 (54/61).

In the cross-validation set model (2) classifies correctly 83.3% and 86.4% of active and inactive compounds in the prediction set, respectively, for a global classification of 85.7%. The percentage of false actives and false inactive are 10.7% (6/56) and 3.6% (2/56), respectively, and one compound was unclassified (1.8%). In Tables 3 and 4 we give the classification of compounds in the external prediction set.

The classification results in the external prediction set were of 85% of good classification because the three active compounds in this set were identified correctly by model (2) and 14 of the 17 inactive were classified correctly. Compounds in the cross validation set as well as in the external prediction set are given in Tables 3 and 4 together with those in the training set.

Structural Interpretation. The structural interpretation of the discriminant function (2) is not an easy task because of the great molecular diversity of the compounds used to derive the model. The TOPS-MODE approach permits one to identify the quantitative contribution of the different fragments or groups to the biological activity under study. It is straightforward to realize that the biological activity of a molecule is the result of the contribution of all its fragments and is not necessarily determined by the influence of only one specific fragment. However, the identification of such possible structural features is interesting in order to orientate the synthesis of novel derivatives by minimizing the groups that contribute negatively to the activity and maximizing

Table 3. Molecular Structures, Experimental Activities in MT4 Assay and Calculated Probabilities of Active According to the TOPS-MODE Model Developed in This Work for Some of the Pyrimidyl Nucleoside Studied in the Training, Cross-Validation and External Data Sets



no.	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	act obs	act calc	prob
1	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{21}$	+	56.7
2	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH_2$	-H	$+^{22}$	+	57.2
3	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	-H	$-^{22}$	U	48.5
4	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-C_2H_5$	$-^{22}$	−	25.6
5[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH(CH_2)_3CH_3$	-H	$-^{23}$	−	1.53
6	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH(CH_2)_3COOCH_3$	-H	$-^{23}$	−	0.22
7	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH(CH_2)_3COOH$	-H	$-^{23}$	−	1.62
8	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH(CH_2)_3NH_2$	-H	$-^{23}$	−	0.74
9	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	-OH	-H	$+^{21}$	+	68.4
10	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{21}$	+	75.2
11	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\beta N_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{24}$	−	1.96
12[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	-OH	$-C_2H_5$	$-^{21}$	−	44.6
13	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	-OH	-Cl	$+^{21}$	+	64.7
14[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	$-NH_2$	-H	$+^{21}$	+	75.2
15	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	$-NH_2$	$-CH_3$	$+^{21}$	+	82.2
16	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	$-NHCH_3$	$-CH_3$	$-^{21}$	U	48.8
17[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	$-SCH_3$	$-CH_3$	$-^{25}$	−	7.81
18[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha NO_2$	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{26}$	+	73.1
19	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	-H	$+^{21}$	+	64.9
20	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{21}$	+	72.7
21	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-SH	$-CH_3$	$+^{21}$	+	70.8
22	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	-Cl	$+^{21}$	+	63.6
23	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	-Br	$+^{21}$	+	62.4
24	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	-I	$+^{21}$	+	60.7
25	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	$-NH_2$	$-CH_3$	$+^{21}$	+	79.8
26	O	C (R)	-CH	-CH	C (S)		-H	-F	-H	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{21,22}$	U	49.0
27[a]	O	C (R)	-CH	-CH	C (S)		-H	-F	-H	-H	$-CH_2OH$	$-NH_2$	$-CH_3$	$-^{22}$	+	58.7
28	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	$-C_2H_5$	$-^{24}$	−	41.3
29	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	-OH	-F	$-^{24}$	+	62.8
30	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	$-SCH_3$	-H	$-^{25}$	−	28.8
31	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha F$	-H	$-CH_2OH$	$-OCH_3$	-H	$-^{22}$	−	24.6
32[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha NH_2$	-H	$-CH_2OH$	$-NHCH_3$	$-CH_3$	$-^{25}$	+	86.7
33[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha CF_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{27}$	−	21.7
34[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha CF_3$	-H	$-CH_2OH$	-OH	-H	$-^{27}$	−	15.6
35[a]	S	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha CH_2OH$	-H	$-CH_2OH$	-OH	-CH=CHBr	$-^{28}$	−	0.06
36[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	-Cl	$-^{24}$	−	42.9
37[a]	O	C (R)	-CH	-CH	C (S)		-H	-OH	$-\alpha CN$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{29}$	+	98.9
38	O	C (R)	-CH	-CH	C (S)		-H	-OH	$-\alpha C{\infty}CH$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{29}$	+	76.8
39	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha NH_2$	-H	$-CH_2OH$	$-SCH_3$	$-CH_3$	$-^{25}$	−	35.6
40	-CH	C (S)	O	-CH	C (S)	-H	-H		-H	$-N_3$	$-CH_2OH$	$-NH_2$	-H	$-^{30}$	+	64.1
41[a]	-CH	C (S)	O	-CH	C (S)	-H	-H		-H	$-NH_2$	$-CH_2OH$	$-NH_2$	-H	$-^{30}$	+	95.9
42	-CH	C (R)	-CH	O	C (R)	-H	-H	-H		$-CH_2N_3$	$-CH_2OH$	-OH	$-CH_3$	$-^{31}$	−	5.33
43	-CH	C (R)	-CH	O	C (R)	-H	-H	-H		$-CH_2N_3$	$-CH_2OH$	-OH	-H	$-^{31}$	−	3.64
44	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha N_3$	-H	$-CH_2CH_2OH$	-OH	$-CH_3$	$-^{32}$	−	14.7
45	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH(OH)CH_2OH$	-OH	$-CH_3$	$-^{32}$	U	48.1
46[a]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH(OH)CH_3$	-OH	$-CH_3$	$-^{32}$	+	95.3
47	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha N_3$	-H	$-CH(OH)CH_2OH$	-OH	$-CH_3$	$-^{32}$	−	14.4
48[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha N_3$	-H	$-CH(OH)CH_3$	-OH	$-CH_3$	$-^{32}$	+	77.9
49	O	C (R)	-CH	C	C (S)		-H	-H	=NOH	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{26}$	−	39.9
50	O	C (R)	-CH	C	C (S)		-H	-H	$=NOCH_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{26}$	−	5.95
51	O	C (R)	CH	C	C (S)		-H	-H	$=NOCOCH_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{26}$	+	64.8
52[a]	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_3$	$+^{21}$	+	56.7
53	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH_2$	-H	$+^{21}$	+	57.2
54	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	$-NH_2$	-Cl	$-^{21}$	+	53.9
55	O	C (R)	C =	C	C (S)		-H	-H	$-CF_3$	-H	$-CH_2OH$	-OH	$-CH_3$	$-^{27}$	−	2.10
56[a]	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	-H	$-^{24}$	U	48.5
57[a]	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	-Cl	$-^{24}$	−	42.9
58	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-H	$+^{21,33}$	−	37.7
59[a]	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NHNH_2$	-H	$-^{34}$	−	23.1
60	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	-imidazol	-H	$-^{34}$	−	39.4
61	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	$-CH_3$	$-^{35}$	−	26.9
62	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-H	$-^{35}$	−	19.3
63	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-H	$-^{35}$	−	14.9
64[a]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-F	$-^{35}$	−	13.7
65	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-Cl	$-^{35}$	−	14.1
66	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-Br	$-^{35}$	−	13.5
67[a]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-I	$-^{35}$	−	12.7
68	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	$-C_2H_5$	$-^{36}$	−	6.30
69	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	$-CF_3$	$-^{36}$	−	2.50

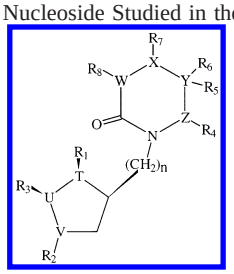
Table 3 (Continued)

no.	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	act obs	act calc	prob
70[a]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	−[36]	−	19.0
71[a]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-Br	−[36]	−	18.8
72[a]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-I	−[36]	−	17.7
73[a]	S	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	$-CH_3$	−[37]	−	12.1
74[a]	S	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-H	−[37]	−	8.20
75	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	-NHOH	-H	−[34]	−	23.1
76	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NH(CH_2)_3CH_3$	-H	−[34]	−	1.55
77	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NHCH_3$	-H	−[34]	−	22.9
78	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NHCH_2C{\infty}CH$	-H	−[34]	−	4.84
79	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NHCH_2CH{=}CH_2$	-H	−[34]	−	4.84
80[a]	O	C (R)	-CH	S	C (S)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	+[38]	+	94.4
81	S	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	+[39]	+	78.4
82[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	-CN	$-CH_2OH$	$-NH_2$	-H	+[40]	+	98.4
83[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-CH{=}CH_2$	$-CH_2OH$	-OH	$-CH_3$	+[41]	+	84.6
84[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-CH{=}CH_2$	$-CH_2OH$	$-NH_2$	-H	+[40]	+	82.3
85[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-CH{=}CHCl$	$-CH_2OH$	$-NH_2$	-H	+[40]	−	16.5
86[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-C{\infty}CH$	$-CH_2OH$	$-NH_2$	-H	+[40]	+	82.3
87[a]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-CH_3$	$-CH_2OH$	$-NH_2$	-H	+[40]	+	97.2
88	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-C_2H_5$	$-CH_2OH$	$-NH_2$	-H	+[40]	+	82.3
89[a]	-CH	N (S)	-CH	-CH	C (R)	-H		-H	-H	-H	$-CH_2OH$	-OH	-H	−[42]	−	5.24
90[a]	-CH	N (S)	-CH	-CH	C (R)	-H		-H	-H	-H	$-CH_2OH$	-OH	$-CH_3$	−[42]	−	6.97
91[a]	-CH	N (S)	-CH	-CH	C (R)	-H		-H	-H	-H	$-CH_2OH$	$-NH_2$	-H	−[42]	−	7.52
92[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	-OH	-H	−[43]	−	23.2
93[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	-OH	-F	−[43]	−	19.8
94[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	-OH	-I	−[43]	−	18.7
95[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	-OH	$-C_2H_5$	−[43]	−	10.3
96[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	-OH	$-CH{=}CHBr$	−[43]	−	0.49
97[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	$-NH_2$	-F	−[43]	−	27.1
98	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	$-NH_2$	-Cl	−[43]	−	27.7
99[a]	$-OCH_2$	C (S)	-CH	-CH	C (R)		-H	-H	-αOH	-H	$-CH_2OH$	$-NH_2$	-I	−[43]	−	25.8
100	O	C (S)	-CH	-CH	C (S)		-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	$-SCH_3$	$-CH_3$	−[25]	−	6.46
101[a]	O	C (S)	-CH	-CH	C (R)		-H	-H	-αF	-H	$-CH_2OH$	$-SCH_3$	-H	−[25]	−	24.6
102	O	C (S)	-CH	S	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-H	+[33,44]	−	7.91
103[a]	S	C (S)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	-OH	$-CH_3$	−[37]	−	34.5
104[a]	S	C (S)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	-OH	-H	−[37]	−	25.7
105[a]	O	C (S)	-CH	S	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	+[38]	+	68.9
106[a]	S	C (S)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-H	+[39]	−	10.7
107	S	C (S)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	+[39]	+	93.5
108	S	C (S)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	$-CH_3$	−[37]	−	8.95
109[a]	S	C (S)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	-OH	-H	−[37]	−	6.05
110[a]	S	C (R)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	-OH	$-CH_3$	−[37]	−	41.8
111[a]	S	C (R)	-CH	O	C (S)		-H	-H		-H	$-CH_2OH$	-OH	-H	−[37]	−	31.9
112[a]	O	C (S)	C =	C	C (R)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2CONH(CH_2)_6NH_2$	−[45]	−	00.0
113[a]	O	C (S)	C =	C	C (R)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2CONH(CH_2)_8NH_2$	−[45]	−	00.0
114[a]	O	C (S)	C =	C	C (R)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2CONH(CH_2)_{10}NH_2$	−[45]	−	00.0
115[a]	O	C (S)	C =	C	C (R)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2CONH(CH_2)_{12}NH_2$	−[45]	−	00.0
116[a]	-CH	C (S)	-CH	-CH	C (R)	-αF	-H	-H	-αF	-H	$-CH_2OH$	-OH	$-CH_3$	−[46]	−	29.9
117	O	C (R)	-CH	-CH	C (R)		-H	-H	$-\alpha OCH_2-$	O	$-CH_2OH$	-OH	$-CH_3$	−[41]	−	45.1
118[b]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-C{\infty}CH$	$-CH_2OH$	-OH	$-CH_3$	+[41]	+	84.6
119[b]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha SCOCH_3$	-H	$-CH_2OH$	-OH	$-CH_3$	−[29]	+	100
120[b]	O	C (R)	-CH	-CH	C (R)		-H	-H	-αOH	$-N_3$	$-CH_2OH$	-OH	$-CH_3$	+[21]	+	61.8
121[b]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\beta N(OH)CH_2Ph$	-H	$-CH_2OH$	-OH	$-CH_3$	−[47]	−	0.02
122[b]	O	C (R)	-CH	-CH	C (S)		-H	-H	$-\alpha NH_2$	-H	$-CH_2OH$	-OH	$-CH_3$	−[48,49]	+	95.6
123[b]	O	C (R)	-CH	-CH	C (S)		-H	-H	-αCN	-H	$-CH_2OH$	-OH	$-CH_3$	−[24]	−	40.8
124[b]	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2NH(CH_2)_6NHCOCF_3$	−[50]	−	0.03
125[b]	O	C (R)	C =	C	C (S)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_2NH-CH_2C{\infty}CH$	−[50]	−	0.80
126[b]	O	C (R)	C =	C	C (S)		-H	-H	-H	$-CF_3$	$-CH_2OH$	-OH	$-CH_3$	−[51]	−	2.66
127[b]	C	C (R)	-CH	O	C (R)	$=CH_2$	-H	-H	-H	-H	$-CH_2OH$	-OH	-H	−[52]	−	7.55
128[b]	C	C (R)	-CH	O	C (R)	$=CH_2$	-H	-H	-H	-H	$-CH_2OH$	$-NH_2$	-H	−[52]	−	10.4
129[b]	-CH	C (S)	-CH	-CH	C (R)	$-CH_2$-OH	-H	-H	-H	-H	-H	-OH	-Cl	−	−	2.25
130[b]	-CH	C (S)	-CH	-CH	C (R)	$-CH_2$-OH	-H	-H	-H	-H	-H	-OH	$-CH_3$	−	−	4.36
131[b]	O	C (R)	-CH	O	C (R)		-H	-H		-H	$-CH_2OH$	$-NH_2$	-F	−[36]	−	19.0
132[b]	-CH	C (S)	-CH	-CH	C (S)	$-\beta$F	-H	-H	$-\alpha N_3$	-H	$-CH_2OH$	-OH	$-CH_3$	−[46]	−	31.8
133[b]	O	C (R)	-CH	-CH	C (R)		-H	-H	-H	-H	$-CH_2OH$	-OH	$-CH_3$	−[24]	+	86.7

[a] Compound in a cross-validation set. [b] Compound in an external prediction set.

those with positive contribution. For instance, model (2) is able to classify correctly compounds for which slight variations in their structures produce changes in the anti-HIV activity. Such changes can be useful for exploring some general structural rules toward the synthesis of novel active compounds. These rules are formulated on the following basis. We first calculate the contribution of a large pool of structural fragments to the anti-HIV activity. Then we make groups of related fragments from which some general trends can be extracted. These groups are completed with some fragments or substructures that can be necessary for formulating a rule. For instance, in formulating the rule (vi) we make a group with tetrahydrofuran, oxathiolane, and dioxolane with configurations 1'R, 4'S and we observed the following order of contribution: tetrahydrofuran > oxathiolane > dioxolane. Finally, we also calculate the contributions of these rings with other configurations, i.e., 1'S, 4'S, and formulate the general rule that is tried to prove from the

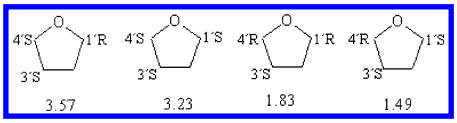
Table 4. Molecular Structures, Experimental Activities in MT4 Assay and Calculated Probabilities of Active According to the TOPS-MODE Model Developed in This Work for Some of the Pyrimidyl Nucleoside Studied in the Training and External Data Sets



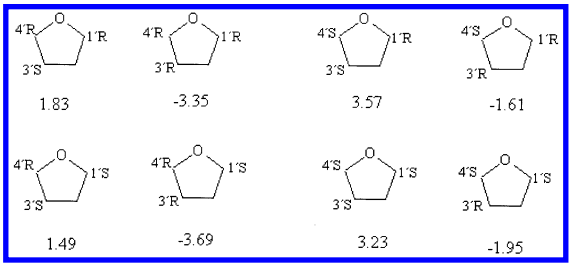
no.	T	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	n	act obs.	act calc	prob
134	O	-CH	S	C=	C	C=	C			$-CH_2-OH$	-H	-H		$-NH_2$	-F	0	−[34]	−	15.4
135	O	-CH	O	C=	C	C=	C			$-CH_2-OH$	-H	-H		$-NH_2$	-F	0	−[34]	−	3.8
136	O	-CH	S	C=	C	C=	C			$-CH_2-OH$	-H	-F		$-NH_2$	-H	0	−[34]	−	36.3
137	O	-CH	-CH	N	C	C=	C		$-\alpha N_3$	$-CH_2-OH$	-H	$-CH_3$		=O	$-CH_2-C_6H_5$	0	−[53,54]	−	0.22
138[a]	O	-CH	-CH	N=	C	C	-CH		$-\alpha N_3$	$-CH_2-$	O-	$-CH_3$	-Br	-OH		0	+[55]	+	95.5
139[a]	-CH	-CH	-CH	N=	C	C=	C	$-CH_2$-OH	-H	-H	-H	-H		-OH		1	−[56]	−	1.33
140[a]	O	-CH	-CH	N=	C	C=	C		$-O-CH_2-$	-CHOH-	-H	-H		$-NH_2$		0	−[57]	−	6.59
141[a]	CH=	C	-CH	N=	C	C=	C	-H	-H	$-CH_2-OH$	-H	$-CH_3$		-OH		0	−[58]	−	13.3

[a] Compund in an external prediction set.

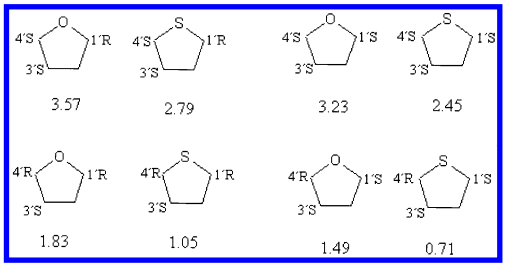
Scheme 1



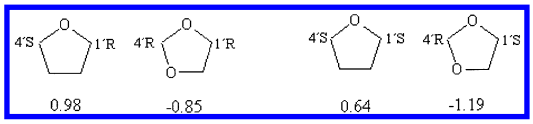
Scheme 2



Scheme 3



Scheme 4



experimental data available. Among these rules we can mention the following:

(i) The preferred configuration of the sugar ring is 1'R, 3'S, 4'S. The following are also welcomed for the development of activity: 1'S, 3'S, 4'S; 1'R, 3'S, 4'R; 1'S, 3'S, 4'R (see Scheme 1). For instance, the configuration of the sugar ring in AZT (compound **10** in Table 3) is 1'R, 3'S, 4'S.

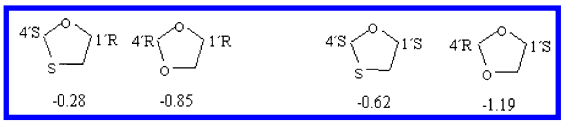
(ii) The change of configuration in 3' from S to R makes inactive the compounds. Thus, this position is critical for the development of anti-HIV activity in this kind of derivatives (see Scheme 2). For instance, the change of configuration in AZT (compound **10** in Table 3) of the group α-N_3 to β-N_3 (compound **11** in Table 3) makes inactive the last compound.

(iii) The substitution of the oxygen atom in the tetrahydrofuran ring by a sulfur atom decreases the probability of activity in the MT4 assay (see Scheme 3). This rule is corroborated by the fact that the thio analogues of AZT and ddI were inactive againts HIV, and the thio analogue of ddC did show only some modest activity in vitro.[3]

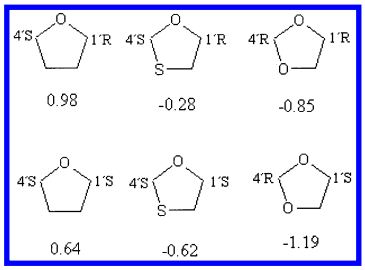
(iv) The change from tetrahydrofuran to dioxolane rings with configurations 1'R, 4'S or 1'S, 4'S transforms the contribution to the activity from positive to negative one (activity greater than 10 μM) in MT4 assay (see Scheme 4). The change of configuration at the position 4' is due to the change of priorities in the atoms after the introduction of the second oxygen atom. An example ratifying this rule is given by compounds **2** and **62** (Table 3) where the introduction of a second oxygen transforms compound **2** (active) into an inactive one (**62**). Despite this, compound **62**, which is named (+)-BCH-204, has potent activity against HIV-1 replication in vitro as reported by Kim et al.[35] and Mansour et al.[34] A possible cause of the decrease in activity after the introduction of the second oxygen can be due to a decrease in the phosphorylation capacity of the dioxolane analogue. For instance, only moderate activity was reported for the (−)-β-D enantiomer of dioxolane-T apparently due to inefficient phosphorylation in some cell lines as evidenced by the antiviral activities obtained with the monophosphate bioisosters.[3]

(v) Oxathiolane rings with configurations 1'R, 4'S show greater probability of activity in MT4 assay to that in which two oxygen atoms are at similar positions (dioxolane ring) and the configuration is 1'R, 4'R. The change of configuration at the position 4' is due to the change of priorities in the atoms after the introduction of the sulfur atom. The same is true for the ring with O, S and configurations 1'S, 4'S compared to that with O, O and configurations 1'S, 4'R (see Scheme 5). For instance, compound **58** in Table 3 (3TC, Epivir) was approved for clinical use in combination with

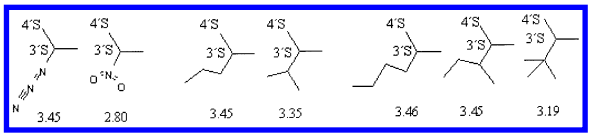
Scheme 5



Scheme 6



Scheme 7



AZT.[3] Compound **80** in Table 3, (−)-FTC, has comparable activity to 3TC.[3] Both compounds are predicted here as active in MT4 assay.

(vi) In general, the contribution to the anti-HIV activity for the rings studied here is the following: tetrahydrofuran > oxathiolane > dioxolane (see Scheme 6). This order coincides with the order of polarity of these rings. For instance, according to PM3 calculations[80] (experimental values are in parentheses when they have been reported in ref 81) the dipole moments of these three rings are as follows: tetrahydrofuran 1.667 D (1.75 D); 1,3-oxathiolane 1.552 D; 1,3-dioxolane 1.256 D (1.47 D). These differences in polarity could be involved in different orders of phosphorylation as commented previously for some case.

(vii) When considering two substituents at position 3' with similar electronic properties, the linear substituent has more probability of active than branched ones (see Scheme 7). For instance, compounds **10** (AZT) and **18** differ only in the substituent at position 3', the first has the group N_3 and the second a group NO_2. Compound **10** (Table 3) has greater probability of active as well as it is about 1000 times more potent as anti-HIV than compound **18**.[82] The main reason for these differences can be found in the interaction of the group at this position with G112 in the HIV−RT. A superior width limit of about 3.7 Å of the 3'-substituent sugar part in the active compound is found.[6] The branching in a substituent at this position increases this width perturbing the interaction with the G112 and decreasing the probability of active as shown in Scheme 7.

(viii) The probability of active in MT4 assay decreases as the size of the substituent at position 5 of the pyrimidine base (R_8) increases. The order is, for instance, CH_3 > Cl > Br > I ≫ C_2H_5 (which is inactive) as observed in the probabilities of active for compounds **20**, **22**, **23**, **24** and **28** (Table 3) which are only different in the substituent R_8. An exception is F that has a probability similar to Cl.

QSAR Model. To develop QSAR models that permit to predict the potency of nucleosides selected with the use of the discriminant model (2) as active we used a series of 2D and 3D connectivity indices as explained in a previous section. These QSAR models were developed for a reduced set of compounds, that is for those having Log1/EC_{50} under 10 when the EC_{50} is expressed in mol/L as in ref 8. The best QSAR model obtained by using this combination of molecular descriptors included two spectral moments of the bond matrix, three 2D connectivity indices and one topographic molecular index, the 3D connectivity index based on atomic charge weighted molecular graph. This model together with its statistical parameters is given below

$$\log 1/EC_{50} = 8.443(\pm 1.198) - 1.355(\pm 0.193)\mu_1 + 0.099(\pm 0.013)\mu_3 + 265.699(\pm 55.974)[{}^5\Omega_C(q)] - 10.956(\pm 1.425)[{}^6\chi_p{}^v] + 9.706(\pm 2.498)[{}^3\chi_C{}^v] - 260.611(\pm 51.267)[{}^5\chi_C{}^v] \quad (3)$$

$$N = 28,\ R = 0.938,\ R^2 = 0.880,\ s = 0.516,\ RMSECV = 0.629$$

where N is the number of compounds included in the model, R is the correlation coefficient, s is the standard deviation of the regression, and *RMSECV* is the mean squared error of the leave-one-out cross-validation. In developing this model six nucleoside derivatives were detected as statistical outliers according to the different tests explained in a previous section. These compounds are those with numbers **3**, **10**, **20**, **54**, **83**, and **118** in Table 3. A close inspection of the molecular descriptors included in the QSAR model showed that several of these indices are strongly interrelated to each other as happens many times when these kinds of molecular descriptors are used in QSAR. In Table 5 we give the correlation coefficients of the six descriptors in model (3).

Table 5. Intercorrelation of the Variables (Molecular Descriptors) Included in the QSAR Model

	μ_1	μ_3	${}^5\Omega_C(q)$	${}^6\chi^v_p$	${}^3\chi^v_C$	${}^5\chi^v_C$
μ_1	1.00	0.90	0.33	0.37	0.26	0.31
μ_3		1.00	0.53	0.55	0.47	0.52
${}^5\Omega_C(q)$			1.00	0.91	0.99	0.99
${}^6\chi^v_p$				1.00	0.88	0.90
${}^3\chi^v_C$					1.00	0.99
${}^5\chi^v_C$						1.00

Table 6. Regression Coefficients in the QSAR Model for Orthogonal Molecular Descriptors

${}^3\chi^v_C$	${}^6\chi^v_p$	μ_3	μ_1	${}^5\chi^v_C$	${}^5\Omega_C(q)$	intercept
1.5156						5.3536
1.5156	−10.4390					5.3536
1.5156	−10.4390	0.0101				5.3536
1.5156	−10.4390	0.0101	−0.7559			5.3536
1.5156	−10.4390	0.0101	−0.7559	−23.0134		5.3536
1.5156	−10.4390	0.0101	−0.7559	−23.0134	265.6990	5.3536

In Table 6 we resume the results of the orthogonalization of molecular descriptors included in model (3). In this case, the last row corresponds to the final model with the orthogonalized molecular descriptors. In Table 6 we can observe the great stability of the regression coefficients after the orthogonalization of molecular descriptors.

The observed and calculated values of Log1/EC_{50} for the 28 compounds studied are given in Table 7 together with

Table 7. Experimental and Predicted Values of the log $1/EC_{50}$ in the MT4 Assay for Compounds Used in the QSAR Model as Well as Residual and Leaves-One-Out Cross-Validation (Delete Residual)

no.[a]	observed value	predicted value	residual	deleted residual
1	5.51	6.39	−0.89	−0.98
2	6.74	6.22	0.53	0.61
12	4.19	4.68	−0.49	−0.67
13	6.14	5.66	0.49	0.52
14	5.12	5.56	−0.44	−0.50
15	5.74	5.56	0.19	0.22
16	4.77	4.88	−0.11	−0.13
17	4.51	4.26	0.25	0.59
18	5.43	5.70	−0.27	−0.31
19	6.62	5.82	0.80	0.91
21	6.00	6.03	−0.03	−0.07
22	6.27	6.18	0.09	0.10
23	5.57	6.43	−0.87	−1.06
24	6.80	6.61	0.19	0.25
25	5.72	6.08	−0.36	−0.40
31	4.34	4.61	−0.27	−0.33
49	7.01	6.48	0.53	0.57
50	4.91	5.10	−0.19	−0.21
51	6.77	6.12	0.65	1.15
52	8.00	7.44	0.56	0.66
53	6.89	7.18	−0.29	−0.36
82	8.92	9.49	−0.57	−1.34
84	8.07	8.03	0.03	0.04
85	5.68	5.82	−0.14	−0.16
86	8.66	8.54	0.11	0.15
87	7.21	7.44	−0.24	−1.04
88	7.89	7.18	0.70	0.82
138	8.10	8.06	0.04	0.34

[a] See Tables 3 and 4 for structures of the compounds according to their numbers.

the residuals and the leave-one-out cross validation residuals (delete residuals). This QSAR model permits one to predict quantitatively the potency of a pyrimidyl derivative in the MT4 assay, which allows one to optimize the activity for compounds selected in the previous step of drug discovery.

CONCLUSIONS

The main conclusion of this work is that we have been able to develop in silico models for the main two steps of drug discovery: lead generation and lead optimization. The first step is fulfilled by using the discriminant model developed with the TOPS-MODE approach. This model permits one to classify pyrimidyl nucleosides as active or inactive ones in the MT4 cell assay. The finding of some rules based on the fragment contributions derived from this approach permits one to direct the synthesis of these compounds to those with higher probability of activity avoiding certain structures with low probability of anti-HIV activity. The second step is carried out when these lead compounds are discovered or selected in the lead generation step. Then, we can generate in silico a series of analogues of these lead compounds and optimize their activities by using the QSAR model developed with topological and topographic descriptors. In this step of lead optimization we are interested in maximizing the biological activity of such active compounds by selecting appropriate substituents at different positions and predicting with the QSAR model the activity of such compounds. In closing, the current study will permit a more rational selection and design of pyrimidyl nucleosides with anti-HIV activity.

ACKNOWLEDGMENT

Thanks are given to the Xunta de Galicia (PR405A 2001/65-0) and Ministerio de Educación y Cultura, Spain (project PM99-0125) for partial financial support.

REFERENCES AND NOTES

(1) Schinazi, R. F. Competitive inhibitors of human immunodeficiency virus reverse transcriptase. *Persp. Drug Discuss. Des*. **1993**, *1*, 151−180.
(2) De Clercq, E. Toward improved anti-HIV chemotherapy: Therapeutic strategies for intervention with HIV infections. *J. Med. Chem*. **1995**, *38*, 2492−2517.
(3) Mansour, T. S.; Storer, R. Antiviral nucleosides. *Curr. Pharm. Des*. **1997**, *3*, 227−264.
(4) el Kouni, M. H. Trends in the design of nucleoside analogues as anti-HIV drugs. *Curr. Pharm. Des*. **2002**, *8*, 581−593.
(5) Yates, P. C. A molecular modeling study of nucleoside analogues as potential anti-AIDS drugs. *Struct. Chem*. **1991**, *2*, 611−619.
(6) Pepe, G.; Meyer, M.; Faury, P.; Graciet, J. C.; Chermann, J. C.; Kraus, J. L. A model allowing the design of modified nucleosides as HIV−RT inhibitors. *Eur. J. Med. Chem*. **1996**, *31*, 775−786.
(7) Dimoglo, A. S.; Gorbachov, M. Y.; Lesnik, T. I.; Saracoglu, M.; Guzel, Y.; Yildrim, I. Investigation of the relationship between chemical structure and anti-HIV-1 activity in a class of nucleoside analogues: electron-topological approach. *Curr. Med. Chem*. **1997**, *4*, 23−34.
(8) Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K.; Hansch, C. Comparative quantitative structure−activity relationships studies on anti-HIV drugs. *Chem. Rev*. **1999**, *99*, 3525−3601.
(9) Mickle, T.; Nair, V. Predictive QSAR analysis of anti-HIV agents. *Drugs Future* **2000**, *25*, 393−400.
(10) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem*. **2001**, *8*, 1573−1588.
(11) Estrada, E. On the topological sub-structural molecular design (TOSS-MODE) in QSPR/QSAR and drug design research. *SAR QSAR Environ. Res*. **2000**, *11*, 55−73.
(12) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 844−849.
(13) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 320−328.
(14) Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 3. Molecules containing cycles. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 23−27.
(15) Estrada, E.; Gutierrez, Y. Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *J. Chromatogr. A* **1999**, *858*, 187−199.
(16) Estrada, E., Peña, A.; García-Domenech, R. Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. *J. Comput.-Aided Mol. Des*. **1998**, *12*, 583−595.
(17) Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L.; De Clercq, E. A novel approach to the rational selection and design of anticancer compounds. *J. Med. Chem*. **2000**, *43*, 1975−1985.
(18) Estrada, E.; Peña, A. In Silico studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem*. **2000**, *8*, 2755−2770.
(19) Estrada, E.; Uriarte, E. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res*. **2001**, *12*, 309−324.
(20) Estrada, E.; Molina, E.; Uriarte, E. Quantitative structure-toxicity relationships using TOPS-MODE. 2. Neurotoxicity of a noncongeneric series of solvents. *SAR QSAR Environ. Res*. **2001**, *12*, 445−459.
(21) De Clercq, E. HIV inhibitors targeted at the reverse transcriptase**.** *AIDS Res. Hum. Retroviruses* **1992**, *8*, 119−134.
(22) De Clercq, E.; Van Aerschot, A.; Herdewijn, P.; Baba, M.; Pauwels, R.; Balzarini, J. Anti-HIV-1 activity of 2′,3′-dideoxynucleoside analogues: structure−activity relationship. *Nucleosides Nucleotides* **1989**, *8*, 659−671.
(23) Dechaux, E.; Pontikis, R.; Monneret, C. Efficient synthesis of N-4 alkyl derivatives of 2′,3′-dideoxycytidine (ddC). *Nucleosides Nucleotides* **1999**, *18*, 1−4.
(24) De Clercq, E. *Perspectives for the chemotherapy of HIV infection: an introduction*; Pharmacochem. Libr.: 1990; Vol. 14 (Des. Anti-AIDS Drugs), pp 1−24.
(25) Abdel-Rahman, A. A.-H.; Abdel-Bary, H. M.; Pedersen, E. B.; Nielsen, C. Synthesis of 4-methylthio analogues of FLT and AZT and their evaluation against HIV. *Arch. Pharm*. **1995**, *328*, 67−70.
(26) Fedorov, I. I.; Kazmina, E. M.; Gurskaya, G. V.; Jasko, M. V.; Zavodnic, V. E.; Balzarini, J.; De Clercq, E.; Faraj, A., Sommadossi, J.-P.; Imbach, J.-L.; Gosselin, G. Novel 3′−C/N-substituted 2′,3′-β-

D-dideoxynucleosides as potential chemotherapeutic agents. 1. Thymidine derivatives: Synthesis, structure, and broad spectrum antiviral properties. *J. Med. Chem.* **1997**, *40*, 486-494.

(27) Lavaire, S.; Plantier-Royon, R.; Portella, C.; De Monte, M.; Kirn, A.; Aubertin, A.-M. 3′-Deoxy-3′−C-trifluoromethyl nucleosides: synthesis and antiviral evaluation. *Nucleosides Nucleotides* **1998**, *17*, 2267−2280.

(28) Ichikawa, E.; Yamamura, S.; Kato, K. Synthesis of 2′,3′-dideoxy-3′−C−(hydroxymethyl)-4′-thiopentofuranosyl nucleosides as potential antiviral agent. *Bioorg. Med. Chem. Lett*. **1999**, *9*, 1113−1114.

(29) Herdewijn, P.; Balzarini, J.; Baba, M.; Pauwels, R.; Van Aerschot, A.; Janssen, G.; De Clercq, E. Synthesis and anti-HIV activity of different sugar-modified pyrimidine and purine nucleosides. *J. Med. Chem.* **1988**, *31*, 2040−2048.

(30) Jeong, L. K.; Lee, Y. A.; Moon, H. R.; Chun, M. W. Synthesis and antiviral activity of apio dideoxy nucleosides with azido or amino substituent. *Nucleosides Nucleotides* **1998**, *17*, 1473−1487.

(31) Jeon, G. S.; Nair, V. New isomeric analogues of anti-HIV active azidonucleosides. *Tetrahedron* **1996**, *52*, 12643−12650.

(32) Hiebl, J.; Zbiral, E.; Balzarini, J.; De Clercq, E. Side-chain derivatives of biologically active nucleosides. 1. Side-chain analogues of 3′-azido-3′-deoxythymidine (AZT). *J. Med. Chem.* **1992**, *35*, 3016−3023.

(33) Schinazi, R. F.; Chu, C. K.; Peck, A.; McMillan, A.; Mathis, R.; Cannon, D.; Jeong, L. S.; Beach, J. W.; Choi, W. B.; Yeola, S. Activities of the four optical isomers of 2′,3′-dideoxy-3′-thiacytidine (BCH-189) against human immunodeficiency virus type 1 in human lymphocytes. *Antimicrob. Agents Chem.* **1992**, *36*, 672−676.

(34) Mansour, T. S.; Evans, C. A.; Siddiqui, M. A.; Charron, M.; Zacharie B.; Nguyen-Ba, N.; Lee, N.; Korba, B. Structure−activity relationship of pyrimidine heterosubstituted nucleoside analogues. *Nucleosides Nucleotides* **1997**, *16*, 993−1001.

(35) Kim, H. O.; Ahn, S. K.; Alves, A. J.; Beach, J. W.; Jeong, L. S.; Choi, B. G,; Van Roey, P.; Schinazi, R. F.; Chu, C. K. Asymmetric synthesis of 1,3-dioxolane-pyrimidine nucleosides and their anti-HIV activity. *J. Med. Chem.* **1992**, *35*, 1987−1995.

(36) Wilson, L. J.; Choi, W. B.; Spurling, T.; Liotta, D. C.; Schinazi, R. F.; Cannon, D.; Painter, G. R.; St. Clair, M.; Furman, P. A. The synthesis and anti-HIV activity of pyrimidine dioxolanyl nucleosides. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 169−74.

(37) Belleau, B.; Brasili, L.; Chan, L.; DiMarco, M. P.; Zacharie, B.; Nguyen Ba, N.; Jenkinson, H. J.; Coates, J. A. V.; Cameron, J. M. A novel class of 1,3-oxathiolane nucleoside analogues having potent anti-HIV activity. Bioorg. *Med. Chem. Lett.* **1993**, *3*, 1723−1728.

(38) Schinazi, R. F Schinazi, R. F.; McMillan, A.; Cannon, D.; Mathis, R.; Lloyd, R. M.; Peck, A.; Sommadossi, J. P.; St Clair, M.; Wilson, J.; Furman, P. A. Selective inhibition of human immunodeficiency viruses by racemates and enantiomers of *cis*-5-fluoro-1-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine. *Antimicrob. Agents Chem.* **1992**, *36*, 2423−31.

(39) Mansour, T. S.; Jin, H.; Wang, W.; Hooker, E. U.; Ashman, C.; Cammack, N.; Salomon, H.; Belmonte, A. R.; Wainberg, M. A. Anti-human immunodeficiency virus and anti-hepatitis-B virus activities and toxicities of the enantiomers of 2′-deoxy-3′-oxa-4′-thiocytidine and their 5-fluoro analogues in vitro. *J. Med. Chem.* **1995**, *38*, 1−4.

(40) Nomura, M.; Shuto, S.; Tanaka, M.; Sasaki, T.; Mori, S.; Shigeta, S.; Matsuda, A. Nucleosides and nucleotides. 185. Synthesis and biological activities of 4′alpha-C-branched-chain sugar pyrimidine nucleosides. *J. Med. Chem.* **1999**, *42*, 2901−2908.

(41) Sugimoto, I.; Shuto, S.; Mori, S.; Shigeta, S.; Matsuda, A. Nucleosides and nucleotides. 183. Synthesis of 4′-alpha-branched thymidines as a new type of antiviral agent. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 385−388.

(42) Mansour, T. S.; Jin, H. Synthesis of (±)-1′-aza-carbocyclic-pyrimidine-2′, 3′-dideoxynucleoside analogues as potential anti-HIV agents. *Bioorg. Med. Chem. Lett.* **1991**, *1*, 757−760.

(43) Verheggen, I.; Van Aerschot, A.; Van Meervelt, L.; Rozenski, J.; Wiebe, L.; Snoeck, R.; Andrei, G.; Balzarini, J.; Claes, P.; et al. Synthesis, biological evaluation, and structure analysis of a series of new 1,5-anhydrohexitol nucleosides. *J. Med. Chem.* **1995**, *38*, 826−835.

(44) Coates, J. A.; Cammack, N.; Jenkinson, H. J.; Jowett, A. J.; Jowett, M. I.; Pearson, B. A.; Penn, C. R.; Rouse, P. L.; Viner, K. C.; Cameron, J. M. (−)-2′-Deoxy-3′-thiacytidine is a potent, highly selective inhibitor of human immunodeficiency virus type 1 and type 2 replication in vitro. *Antimicrob. Agents Chem.* **1992**, *36*, 733−739.

(45) Camara, F.; Ciurea, A.; Delbederi, Z.; Fossey, C.; Fontaine, G.; Gavriliu, D.; Jouenne, J.; Laduree, D.; Aubertin, A. M.; Kirn, A. Synthesis of L-analogues of 1-(2′, 3′-dideoxy-*β*-D-glyceropent-2-enofuranosyl)thymine. *Pharm. Pharmacol. Commun.* **1999**, *5*, 225−231.

(46) Highcock, R. M.; Hilpert, H.; Myers, P. L.; Roberts, S. M.; Storer, R. The potential for using carbocyclic nucleosides for the treatment of AIDS. Part 1. Preparation of some analogues for azidothymidine (AZT). *J. Chem. Soc., Perkin Trans. 1* **1991**, 1127−1134.

(47) Tronchet, J. M. J.; Zsely, M.; Capek, K.; Del Villedon de Naide, F. Synthesis and anti-HIV activity of 3′-deoxy-3′-(hydroxyamino) analogues of nucleosides. *Bioorg. Med. Chem. Lett.* **1992**, *2*, 1723−178.

(48) Chu, C. K.; Schinazi, R. F.; Ahn, M. K.; Ullas, G. V.; Gu, Z. P. Structure−activity relationships of pyrimidine nucleosides as antiviral agents for human immunodeficiency virus type 1 in peripheral blood mononuclear cells. *J. Med. Chem.* **1989**, *32*, 612−617.

(49) Kovacs, T.; Parkanyi, L.; Pelczer, I.; Cervantes-Lee, F.; Pannell, K. H.; Torrence, P. F. Solid-state and solution conformation of 3′-amino-3′-deoxythymidine, precursor to a noncompetitive inhibitor of HIV-1 reverse transcriptase. *J. Med. Chem.* **1991**, *34*, 2595−2600.

(50) Gavriliu, D.; Fossey, C.; Fontaine, G.; Benzaria, S.; Ciurea, A.; Delbederi, Z.; Lelong, B.; Laduree, D.; Aubertin, A. M.; Kirn, A. Synthesis and antiviral activity of C-5 substituted analogues of d4T bearing methylamino- or methyldiamino-linker arms. *Nucleosides, Nucleotides Nucleic Acids* **2000**, *19*, 1017−1031.

(51) Kozak, J.; Johnson, C. R. Synthesis of 4′-trifluoromethyl nucleoside analogues. *Nucleosides Nucleotides* **1998**, *17*(12), 2221−2239.

(52) Jeong, L. S.; Yoo, S. J. Synthesis and antiviral activity of novel isodideoxy nucleosides with exocyclic methylene. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 847−852.

(53) Adams, D. R.; Perez, C.; Maillard, M.; Florent, J. C.; Evers, M.; Henin, Y.; Litvak, S.; Litvak, L.; Monneret, C.; Grierson, D. S. Preparation and anti-HIV activity of N-3-substituted thymidine nucleoside analogues. *J. Med. Chem.* **1997**, *40*, 1550−1558.

(54) Kitade, Y.; Suzuki, A.; Hirota, K.; Maki, Y.; Nakane, H.; Ono, K.; Baba, M.; Shigeta, S. Synthesis and anti-human immunodeficiency virus type 1 (HIV-1) activity of 3-substituted derivatives of 3′-azido-3′-deoxythymidine (AZT), and inhibition of HIV-1 reverse transcriptase by their 5′-triphosphates. *Chem. Pharm. Bull.* **1992**, *40*, 920−924.

(55) Motura, M.; Salomon, H.; Moroni, G.; Wainberg, M.; Brinon, M. C. Synthesis, lipophilicity and anti-HIV activity of a new brominated analogue of zidovudine. *Nucleosides Nucleotides* **1999**, *18*, 337−351.

(56) Santana, L.; Teijeira, M.; Uriarte, E.; Balzarini, J.; De Clercq, E. Synthesis and antiviral activity of 1,2-carbonucleosides. *Nucleosides Nucleotides* **1995**, *14*, 521−523.

(57) Molas, M. P.; Matheu, M. I.; Castillon, S.; Isac-Garcia, J.; Hernandez-Mateo, F.; Calvo-Flores, F. G.; Santoyo-Gonzalez, F. Synthesis of 3,6-anhydro sugars from cyclic sulfites and sulfates and their applications in the preparation of bicyclo nucleoside analogues of ddC and ddA. *Tetrahedron* **1999**, *55*, 14649−14664.

(58) Maag, H.; Rydzewski, R. M. An allylic azide route to 4′-azido carbocyclic nucleosides. Synthesis of (±)−(1′a, 2′a, 3′b)- and (±)−(1′a, 2′b, 3′b)-1-[1-azido-2-hydroxy-1-(hydroxymethyl)-3-cyclopentyl]-thymine. *J. Org. Chem.* **1992**, *57*, 5823−5831.

(59) Ferguson, L. N. *The modern structural theory of organic chemistry*; Englewoods Cliffs: Prentice-Hall: 1963.

(60) Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(61) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(62) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, 1986.

(63) Hall, L. H.; Kier, L. B. Molecular connectivity chi indices for database analysis and structure−property modeling. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, The Netherlands, 1999; pp 307−360.

(64) Estrada, E. Edge adjacency relationships and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 31−33.

(65) Estrada, E. Guevara, N. Gutman, I. Extensión of edge connectivity index. Relationships to line graph indices and QSPR applications. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 428−431.

(66) Estrada, E.; Rodriguez, L. Edge-connectivity indices in QSPR/QSAR studies. 1. Comparison to other topological indices in QSPR studies. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 1037−1041.

(67) Estrada, E.; Montero, L. A. Bond order weighted graphs in molecules as structure−property indices. *Mol. Eng*. **1993**, *2*, 363−373.

(68) Estrada, E.; Ramírez, A. Edge adjacency relationships and molecular topographic descriptors. Definition and QSAR applications. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 837−843.

(69) Estrada, E. Three-dimensional molecular descriptors based on electron charge density weighted graphs. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 708−713.

(70) Estrada, E.; Molina, E. 3D connectivity indices in QSPR/QSAR studies. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 791−797.
(71) Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985**, *107*, 3902−3909.
(72) Stewart, J. J. P. MOPAC 6.0, *Quantum Chemistry Program Exchange*; Indiana University, Bloomington, IN, Program 455.
(73) Rodriguez, L.; Estrada, E. MODEST (MOlecular DESign Tool) for Windows. Version 3.0; 1999−2000.
(74) STATISTICA 6.0; StatSoft Inc.: Tulsa, U.S.A., 2002.
(75) Randić, M. Resolution of ambiguities in structure−property studies by us of orthogonal descriptors. *J. Chem. Inf. Comput. Sci*. **1991**, *31*, 311−320.
(76) Randić, M. Orthogonal molecular descriptors. *New J. Chem*. **1991**, *15*, 517−525.
(77) Randić, M. Correlation of enthalpy of octanes with orthogonal connectivity indices. *J. Mol. Struct.* (*Theochem*) **1991**, *233*, 45−59.
(78) Luèić, B.; Nikolić, S.; Trinajstić, N.; Jurić, D. The structure−property models can be improved using the orthogonalized descriptors. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 532−538.
(79) Klein, D. J.; Randić, M.; Babić, D.; Luèić, B.; Nikolić, S.; Trinajstić, N. Hierarchical orthogonalization of descriptors. *Int. J. Quantum Chem.* **1997**, *63*, 215−222.
(80) Stewart, J. J. P. Optimization parameters for semiempirical methods. I. Method. *J. Comput. Chem*. **1989**, *10*, 209−220.
(81) MacClellan, A. L. *Tables of experimental dipole moments*; Rahara, 1989.
(82) Kotra, L. P.; Xiang, Y.; Newton, M. G.; Schinazi, R. F.; Cheng, Y.-C.; Chu, C. K. Structure−activity relationships of 2-'deoxy-2',2-'difluoro-L-erythro-pentofuranosyl nucleosides. *J. Med. Chem*. **1997**, *40*, 3635−3644.

CI0255331