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

Santiago Vilar,*,† Ernesto Estrada,‡ Eugenio Uriarte,† Lourdes Santana,† and Yaquelin Gutierrez‡

Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela, Santiago de Compostela 15782, Spain, and Complex Systems Research Group, X-ray Unit, RIAIDT, Edificio CACTUS, University of Santiago de Compostela, Santiago de Compostela 15782, Spain

Received November 8, 2004

The TOPological Substructural MOlecular DEsign (TOPS-MODE) approach has been used to predict the anti-HIV activity in MT-4 assays (Estrada et al., 2002) of a diverse range of purine-based nucleosides. A database of 206 nucleosides has been selected from the literature and a theoretical virtual screening model has been developed. The model is able of discriminating between compounds that have anti-HIV activity and those that do not, with a good classification level of 85% in the training and 82.8% in the cross-validation series. On the basis of the information generated by the model, the correct classification of practically 80% of compounds from an external prediction set has been achieved using the theoretical model. Furthermore, the contribution of a range of molecular fragments to the pharmacological action has been calculated and this could provide a powerful tool in the design of nucleoside analogues that show activity against the HIV. Finally, a QSAR model has been developed that allows quantitative data to be obtained regarding the pharmacological potency shown by this type of compound.

INTRODUCTION

The search for new compounds with a given biological activity, or the modification of these compounds to optimize their activity or reduce their toxicity, requires enormous effort in terms of manpower and cost. This effort arises due to the large number of compounds that need to be synthesized and subsequently evaluated from a biological point of view. However, there are a number of technological contributions that represent an advance in the design of novel pharmacological agents, thus reducing the high level of research effort. For this reason the pharmaceutical industry has shown great interest in theoretical methods that enable the rational design of pharmaceutical agents. A number of studies and computerized molecular design techniques have been reported in the literature.[1−3] These techniques are based on the relationship between the molecular structure of compounds and their activity.

The graph-theoretical approach has proven very useful to describe molecular structure and has greatly helped in the development of predictive mathematical models.[4] Indeed, this approach has recently led to the development of a method for rational molecular design called TOPological Substructural MOlecular DEsign (TOPS-MODE). It allows the development of quantitative and qualitative theoretical models that predict the properties of different molecular systems.[5,6] Since the discovery of the agent that causes acquired immunodeficiency syndrome (AIDS), an antiretroviral therapy has evolved aimed at different therapeutic targets.[7,8] One of these is the inhibition of the reverse transcriptase enzyme that avoids the synthesis of viral DNA. This is the mode of action of the various nucleoside analogues that show anti-HIV activity, such as 3′-azido-2′,3′-dideoxythymidine (AZT).[9−12]

In the work described here the TOPS-MODE approach is used to derive a discriminant equation capable of differentiating between purine based nucleosides that are active or inactive against the AIDS virus.

METHODOLOGY

The TOPS-MODE Approach. The TOPS-MODE approach was used in this study with the aim of carrying out virtual screening and the rational design of anti-HIV compounds.

The TOPS-MODE approach is based on the computational calculation of topological indices called spectral moments[13,14] of the edge-adjacency matrix. The spectral moments of the matrix E are defined as

$$\mu_k = \mathrm{tr}(\mathrm{E}^k)$$

The *k*th spectral moment of E is defined as the trace, i.e., the sum of the diagonal entries of the *k*th power of the bond matrix. The edge-adjacency matrix is a symmetrical matrix consisting of ones and zeros depending on whether the bonds are adjacent or not. When the molecular graph contains a heteroatom, it is necessary to differentiate the carbon−heteroatom bonds from the carbon−carbon bonds. This difference is indicated using a series of values that weight the graph. These values or weights refer to different properties of the bond. In this study the dipole moment of the bond is the characteristic used to weight the graph, and the bond dipole values used are shown in Table 1.[15]

* Corresponding author fax: +34-981-594912; e-mail: qosanti@yahoo.es.
† Department of Organic Chemistry, University of Santiago de Compostela.
‡ Edificio CACTUS, University of Santiago de Compostela.

10.1021/ci049662o CCC: $30.25
Published on Web 03/08/2005

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 15.

Database. The initial database developed using the classification model must be as complete and varied as possible in order that the information gives rise to the generation of a high quality discriminant equation. For this reason a range of compounds was included, and these covered significant structural changes in the purine base (see Tables 3−8).

The present study began using a database of 97 molecules, 26 of which show anti-HIV activity in MT-4 assays and 71 of which do not show such activity. This database was divided into two series, a training series and a cross-validation series. The training series consisted of 62 compounds (21 active and 41 inactive), while the cross-validation series contained 35 compounds (5 active and 30 inactive). With the aim of demonstrating the predictive capability of the model, the anti-HIV activities of a series of nucleoside analogues was evaluated. This external prediction series consisted of 109 compounds that were never used in the training or validation stages.

A compound is considered active in this study when it shows an effective concentration $EC_{50} \leq 10\ \mu M$ in MT-4 assays. When the EC_{50} value is greater than 10 μM, the compound is considered inactive or not interesting for the purposes of this work. The aim of this selection is our main interest in the search of highly active compounds only. Consequently, those compounds with $EC_{50} > 10\ \mu M$ are not considered as potentially interesting to be screened for anti-HIV activity in this work. Obviously, this criterion can exclude other compounds which could be potentially active when tested, a risk that is inherent to any virtual screening approach.

Discriminant Analysis. Classification Function. Having calculated the 15 first spectral moments for each compound, a linear discriminant analysis was performed using the STATISTICA package[16] to develop a classification function capable of differentiating active compounds from inactive ones. In eq 1 the activity, or any given biological property, is expressed as a function of the spectral moments generated in each case. In the classification function

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + \ldots\ldots\ldots + a_k\mu_k + b \qquad (1)$$

P is the biological property, μ_k is the kth spectral moment and a_k are the coefficients obtained in the classification function.

The statistical parameters that define the quality of the model are the Wilks' λ and the Mahalanobis distance. The discriminant function was obtained by using the *stepwise* method with a minimum tolerance value of 0.01. The a posteriori probability calculated from the Mahalanobis distance was used for the classification of compounds as active or inactive.

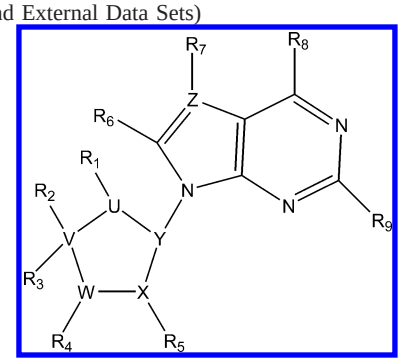
The Use of Variables that Indicate the Chirality of Carbon Atoms. In many cases the stereochemistry of a given compound is closely related to its pharmacological properties. The TOPS-MODE approach is a bidimensional topological approach that does not consider the stereochemical aspects of a molecule, which represents a limitation in terms of differentiating between different enantiomers. This problem was addressed using a series of variables that define the chirality of various chiral centers, which could have great significance for the anti-HIV activity of a given compound. These variables have a value of 1 if the center has an R configuration, −1 if it is S or 0 in cases where it is achiral. The variables I_1, I_2 and I_3 denote chiral centers in positions 1′, 3′ and 4′ or 1′, 2′ and 3′ depending on the size of the ring as shown in Chart 1.

QSAR Model. The linear regression analysis enables a linear relationship to be established between different mo-

Table 2. Topological, Topographic and Quantum Chemical Descriptors Calculated for QSAR Model

symbol	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 $h = 4-6$
${}^h\Omega^c{}_p(q)$	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 $h = 3-6$
${}^h\Omega^c{}_{pC}(q)$	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 $h = 4-6$
E_{LUMO}	LUMO energy
E_{HOMO}	HOMO energy

Table 3. Molecular Structures, Experimental and Calculated Anti-HIV Activities and Calculated Probabilities for Selected Compounds in the Statistical Analysis (Training, Cross-Validation and External Data Sets)



no.	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	act obs[d]	act calc[d]	prob
1	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		NH_2	H	+[34]	+	63.80
2	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		OH	H	+[34]	+	64.42
3	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		OH	NH_2	+[34]	+	85.90
4	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		NH_2	NH_2	+[34]	+	85.30
5	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		OH	OH	+[35]	+	87.12
6	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OH	NH_2	+[34]	+	81.45
7[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NH_2	NH_2	+[34]	+	79.73
8	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NH_2	H	+[36]	U	52.41
9[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	βN_3	H	H		NH_2	H	−[36]	−	17.27
10	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NH_2	H	−[36,37]	−	22.09
11	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	H	H		NH_2	H	−[34]	−	38.64
12	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	βF	H	H		NH_2	H	−[36]	−	10.39
13	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	H	H		OH	NH_2	+[34]	+	69.28
14	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	H	H		NH_2	NH_2	+[34]	+	67.26
15[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	F	H		NH_2	H	−[34,36]	−	12.05
16	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	F	H		NH_2	NH_2	−[36]	−	31.01
17[a]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	H	H		OH	H	−[36]	−	38.26
18[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	OH	H		NH_2	H	−[38]	−	14.26
19[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	Cl	H		NH_2	H	−[38]	−	12.49
20[a]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	βOH	H	H		NH_2	H	−[38]	−	13.54
21	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	βOH	H	H		NH_2	NH_2	−[38]	−	33.53
22	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	βOH	H	H		OH	NH_2	−[38]	−	33.15
23[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	NH_2	H		NH_2	H	−[39]	−	17.96
24	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αNH_2	H	H		NH_2	H	−[39]	−	45.22
25[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	βNH_2	H	H		NH_2	H	−[39]	−	13.13
26	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		H	H	−[35]	−	44.95
27	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OH	H	−[36]	+	54.20
28	O	C(R)	CH	CH	CH(R)	N		CH_2OH	CF_3	αOH	H	H		NH_2	H	−[40]	−	47.29
29	O	C(R)	CH	CH	CH(R)	N		CH_2OH	CF_3	αOH	OH	H		OCH_3	H	−[40]	−	27.89
30[a]	CH	C(S)	CH	O	CH(S)	N	H	CH_2OH	N_3	H		H		NH_2	H	−[41]	−	2.50
31	CH	C(S)	CH	O	CH(S)	N	H	CH_2OH	NH_2	H		H		NH_2	H	−[41]	−	4.64
32	CH	C(R)	O	CH	CH(R)	N	H	CH_2OH	CH_2N_3		H	H		NH_2	H	−[42]	−	17.57
33	O	C(R)	O	CH	CH(R)	N		CH_2OH	H		H	H		OH	NH_2	+[43]	+	78.51
34[a]	O	C(R)	O	CH	CH(R)	N		CH_2OH	H		H	H		NH_2	NH_2	+[43]	+	79.12
35	S	C(R)	O	CH	CH(R)	N		CH_2OH	H		H	H		NH_2	H	+[44]	+	69.32
36	O	C(R)	O	CH	CH(R)	N		CH_2OH	H		H	H		NH_2	H	+[45]	+	53.92
37[a]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		OH	NH_2	+[34,46]	+	59.68
38	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NH-*c*-Pr	NH_2	+[46]	+	85.87
39[a]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		OH	H	−[46]	−	32.66
40[a]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NH-*c*-Pr	H	−[46]	+	57.14
41	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NH-*i*-Pr	NH_2	−[46]	+	59.76
42	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		OCH_3	NH_2	+[46]	+	59.94
43	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NC_3H_6	NH_2	+[46]	U	50.97
44	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		$SCH_2CH{=}CH_2$	NH_2	+[46]	+	61.60
45	O	C(S)	C =	C	CH(R)	N		CH_2OH	H	H	H	H		NH_2	H	−[34]	+	63.80
46[a]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		O-*n*-Bu	NH_2	+[46]	+	54.77
47[a]	O	C(S)	C =	C	CH(R)	N		CH_2OH	H	H	H	H		OH	NH_2	−[36]	+	85.90
48[a]	O	C(S)	C =	C	CH(R)	N		CH_2OH	H	H	H	H		NH_2	NH_2	−[36]	+	85.30
49[a]	O	C(S)	CH−Y		CH(R)	N		CH_2OH	H	αCH_2OH		H		NH_2	H	+[34]	−	6.80
50	O	C(S)	CH−Y		CH(R)	N		CH_2OH	H	αCH_2OH		H		OH	H	+[34]	−	7.23
51[a]	O	C(S)	CH−Y		CH(R)	N		CH_2OH	H	αCH_2OH		H		OH	NH_2	−[34]	−	19.58
52	O	C(S)	CH−Y		CH(R)	N		CH_2OH	H	αCH_2OH		H		NH_2	NH_2	−[34]	−	18.86
53	O	C(S)	CH−Y		CH(R)	N		CH_2OH	H	αCH_2OH		H		OH	OH	−[35]	−	22.17
54[a]	OCH_2	C(R)	CH	CH	CH(S)	N		CH_2OH	H	αOH	H	H		OH	H	−[47]	−	3.15
55	OCH_2	C(R)	CH	CH	CH(S)	N		CH_2OH	H	αOH	H	H		NH_2	NH_2	−[47]	−	8.57
56	OCH_2	C(R)	CH	CH	CH(S)	N		CH_2OH	H	αOH	H	H		H	NH_2	−[47]	−	7.45
57[a]	O	C(S)	CH	CH	CH(R)	C		CH_2OH	H	H	H	H	H	NH_2	H	−[48]	+	81.32
58	O	C(S)	CH	CH	CH(R)	C		CH_2OH	H	H	H	H	$CONH_2$	NH_2	H	−[48]	−	46.42
59[a]	O	C(S)	C =	C	CH(R)	C		CH_2OH	H	H	H	H	H	NH_2	H	−[48]	+	81.32
60[a]	O	C(S)	C =	C	CH(R)	C		CH_2OH	H	H	H	H	CN	NH_2	H	−[48]	−	0.82
61[a]	O	C(S)	C =	C	CH(R)	C		CH_2OH	H	H	H	H	$CONH_2$	NH_2	H	−[48]	−	46.42
62	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	Br		NH_2	H	−[36]	−	32.75
63[a]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	Br		NH_2	H	−[36]	−	25.45
64[a]	O	C(S)	O	CH	CH(S)	N		CH_2OH	H		H	H		OH	NH_2	−[45]	−	9.68
65	S	C(S)	O	CH	CH(S)	N		CH_2OH	H		H	H		NH_2	H	−[44]	−	6.80

Table 3. (Continued)

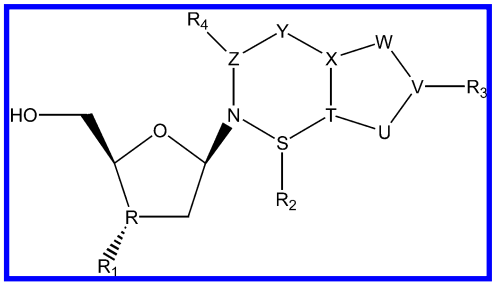
no.	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	act obs[d]	act calc[d]	prob
66	S	C(S)	O	CH	CH(S)	N		CH_2OH	H		H	H		OH	NH_2	−[44]	−	10.55
67	O	C(S)	O	CH	CH(S)	N		CH_2OH	H		H	H		NH_2	H	−[45]	−	3.99
68	CH	C(R)	C =	C	CH(S)	N	CH_2OH	CH_2OH	H	H	H	H		OH	NH_2	−[49]	−	2.88
69	CH	C(S)	C =	C	CH(R)	N	CH_2OH	CH_2OH	H	H	H	H		NH_2	H	+[49,50]	−	16.47
70	CH	C(S)	C =	C	CH(R)	N	CH_2OH	CH_2OH	H	H	H	H		OH	NH_2	−[49]	−	44.58
71[a]	C =	C	CH	CH	CH(R)	N	H	CH_2SCH_3		αOH	OH	H		NH_2	H	−[51]	−	0.00
72[a]	C =	C	CH	CH	CH(R)	N	H	CH_2OCH_3		αOH	OH	H		NH_2	H	−[51]	−	2.96
73	C =	C	CH	CH	CH(R)	N	H	$CH{=}CH_2$		αOH	OH	H		NH_2	H	−[51]	−	9.48
74	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	βO−X		H		NH_2	H	−[38]	−	4.69
75	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αO−X		H		NH_2	H	−[38]	−	0.88
76[a]	C	C(R)	O	CH	CH(R)	N	$=CH_2$	CH_2OH	H		H	H		NH_2	H	−[52]	−	29.01
77[b]	O	C(S)	CH	CH	CH(R)	N		H	CH(OH)-	CH_2-Oβ-	H	H		NH_2	H	−[53]	−	3.61
78[b]	CH	C(S)	CH	CH	N(R)	N	CH_2OH	CH_2OH	H	H	H	H		NH_2	H	−[54]	−	20.29
79[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	OH	H		OH	NH_2	−[34]	−	6.31
80[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	OH	H		NH_2	NH_2	−[34]	−	6.25
81[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	Br	H		NH_2	H	−[38]	−	11.55
82[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αCH_2OH	H	H		OH	NH_2	−[55]	−	30.42
83[b]	S	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		NHMe	H	−[56]	+	80.10
84[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		H	NH_2	−[35]	+	82.88
85[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		SH	NH_2	−[35]	−	5.11
86[b]	O	C(R)	CH	C−F	CH(S)	N		CH_2OH	H	H	F	H		NH_2	H	−[57]	−	0.02
87[b]	O	C(S)	CH	CH	CH(R)	N		CH_2I	H	$\alpha OC(CH_3)_2$-	O-	H		NH_2	H	−[58]	−	1.36
88[b]	C	C(R)	O	CH	CH(R)	N	$=CH_2$	CH_2OH	H		H	H		OH	H	−[52]	−	29.97
89[b]	CH	C	CH−Y		CH	N	H	CH_2OH	CH_2OH	H		H		NH_2	H	−[59]	−	10.11
90[b]		C(R)	C =	C	CH(S)	N		CH_2OH	H	H	H	H		OH	H	−[60]	−	3.43
91[b]	CH	C(S)	CH	CH	CH(S)	N	CH_2−V	CH_2OH		αCH_2OH	H	H		NH_2	H	−[61]	−	0.03
92[b]	CH=CH	C(S)	CH	CH	CH(S)	N		CH_2OH	H	αCH_2OH	H	H		NH_2	H	−[50]	−	1.95
93[b]	S	C(S)	CH	CH	CH(R)	N		CH_2OH	H	βOH	F	H		NH_2	H	−[63]	+	75.19
94[b]	CH	C(R)	C =	C	CH(R)	N	H	$(CH_2)_2OH$	H	H	H	H		OH	NH_2	−[64]	−	46.04
95[b]	S	C(R)	O	CH	CH(R)	N		CH_2OH	H		H	H		OH	NH_2	−[44]	+	84.42
96[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	H	H		NHMe	H	+[34]	+	68.90
97[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αOH	H	H		NH_2	H	+[65]	−	45.58
98[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αOH	H	H		OH	NH_2	+[65]	+	73.00
99[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OMe	NH_2	+[66]	+	82.14
100[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OC_2H_5	NH_2	−[66]	+	81.93
101[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OC_3H_7	NH_2	+[66]	+	82.65
102[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		$OCHMe_2$	NH_2	−[66]	+	91.24
103[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OC_4H_9	NH_2	−[66]	+	81.15
104[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OC_6H_5	NH_2	−[66]	+	96.63
105[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		OBz	NH_2	+[66]	+	94.85
106[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHMe	NH_2	−[66]	+	87.09
107[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHC_2H_5	NH_2	−[66]	+	85.44
108[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHC_3H_7	NH_2	−[66]	+	84.77
109[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHC_4H_9	NH_2	−[66]	+	84.78
110[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHC_6H_5	NH_2	−[66]	+	94.97
111[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		$NHCH_2Bz$	NH_2	−[66]	+	95.53
112[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NMe_2	NH_2	−[66]	+	86.94
113[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		N(Me)Et	NH_2	−[66]	+	90.67
114[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NC_3H_6	NH_2	+[66]	+	80.48
115[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NC_4H_8	NH_2	+[66]	+	93.27
116[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		Cl	NH_2	+[66]	+	80.15
117[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NH-*c*-Pr	NH_2	−[67]	+	96.47
118[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NH-*c*-Bu	NH_2	−[67]	+	88.60
119[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		NHBz	NH_2	−[67]	+	95.47
120[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		N(Me)-*c*-Pr	NH_2	+[67]	+	96.35
121[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		$N(Et)_2$	NH_2	−[67]	+	93.87
122[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		CH_3	NH_2	−[67]	+	79.50
123[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	αN_3	H	H		CN	NH_2	−[67]	−	1.64
124[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	H	αF	H	H		NHMe	H	−[68]	−	45.75
125[b]	O	C(R)	CH	CH	CH(R)	C		CH_2OH	H	αF	H	H	H	NH_2	H	−[68]	+	64.57
126[b]	O	C(S)	CH	CH	CH(S)	N		CH_2OH	H	βF	H	H		NH_2	H	−[68]	−	0.36
127[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	Me		NH_2	H	−[69]	−	0.99
128[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	Cl		NH_2	H	−[69]	−	0.56
129[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	Br		NH_2	H	−[69]	−	0.57
130[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	I		NH_2	H	−[69]	−	1.20
131[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	NHMe		NH_2	H	−[69]	−	0.02
132[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	$CONMe_2$		NH_2	H	−[69]	−	0.53
133[b]	CH	C(R)	CH	CH	CH(S)	N	H	CH_2OH	H	H	H	Ph		NH_2	H	−[69]	−	0.20
134[b]	CH	C(S)	C =	C	CH(S)	N	CH_2OH	CH_2OH	H	CH_2OH	H	H		NH_2	H	−[70]	−	0.22
135[b]	CH	C =	C	CH	CH(S)	N	CH_2OH	CH_2OH		CH_2OH	H	H		NH_2	H	−[70]	−	0.07
136[b]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NH_2	NH_2	+[46]	+	60.01
137[b]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		Cl	NH_2	+[46]	+	59.65
138[b]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		NH_2	H	−[46]	−	27.99
139[b]	CH	C(S)	C =	C	CH(R)	N	H	CH_2OH	H	H	H	H		N(Me)-*c*-C_3H_5	NH_2	+[46]	+	84.00

Table 3. (Continued)

no.	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	act obs[d]	act calc[d]	prob
140[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NHMe	H	−[71]	−	30.71
141[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NMe_2	H	−[71]	−	24.85
142[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NH-*c*-propyl	H	−[71]	+	58.39
143[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NH-*c*-pentyl	H	−[71]	−	36.06
144[b]	O	C(S)	CH	CH	CH(R)	N		CH_2OH	H	H	N_3	H		NH-*c*-hexyl	H	−[71]	U	49.02
145[b]	CH	C(R)	CH	CH	CH(S)	C	H	CH_2OH	H	H	H	H	CN	NH_2	H	−[72]	−	0.01
146[b]	CH	C(R)	CH	CH	CH(S)	C	H	CH_2OH	H	H	H	H	$CONH_2$	NH_2	H	−[72]	−	0.96
147[b]	CH	C(R)	CH	CH	CH(S)	C	H	CH_2OH	H	H	H	H	$C(NH_2)=$	=N−N(Me)-	H	−[72]	−	1.03
148[b]	CH	C(R)	C =	C	CH(S)	N	CH_2−V	CH_2OH		H	H	H		OH	NH_2	−[73]	−	1.04
149[b]	O	C(R)	C−CF_3	CH	CH(R)	N		CH_2OH	H	αOH	OH	H		NH_2	H	−[74]	−	5.12
150[b]	O	C(R)	C−CF_3	CH	CH(R)	N		CH_2OH	H	αOH	OH	H		OH	NH_2	−[74]	−	31.15
151[b]	C	C(S)	O	CH	CH(S)	N	$(CH_2)_2$−U	CH_2OH	H		H	H		NH_2	H	−[75]	−	0.03
152[b]	C	C(S)	O	CH	CH(S)	*c*	$(CH_2)_2$−U	CH_2OH	H		H	*c*	*c*	*c*	*c*	−[75]	−	0.05
153[b]	CH	C(S)	O	CH	CH(S)	N	CH_3	CH_2OH	H		H	H		NH_2	H	−[75]	−	1.26
154[b]	CH	C	CH	CH	CH(S)	N	CH_2OH	H	H	H	H	H		Cl	H	−[31]	−	1.85
155[b]	CH	C	CH	CH	CH(R)	N	CH_2OH	H	H	H	H	H		Cl	H	−[31]	−	28.47
156[b]	CH	C	CH	CH	CH(S)	N	CH_2OH	H	H	H	H	H		NH_2	H	−[31]	−	2.26
157[b]	CH	C	CH	CH	CH(R)	N	CH_2OH	H	H	H	H	H		NH_2	H	−[31]	−	32.75
158[b]	CH	C	CH	CH	CH(S)	N	CH_2OH	H	H	H	H	H		OH	H	−[31]	−	2.30
159[b]	CH	C	CH	CH	CH(R)	N	CH_2OH	H	H	H	H	H		OH	H	−[31]	−	33.25
160[b]	O	C(R)	CH	CH	CH(R)	N		CH_2OH	CF_3	αOH	H	H		OH	H	−[40]	−	41.31

[a] Compound in a cross-validation set. [b] Compound in a external prediction set. [c] The base is adenyn-3-yl. [d] (U) unclassified compounds; (+) active compounds; (−) inactive compounds.

Table 4. Molecular Structures and Anti-HIV Activities (Experimental and Calculated) for Selected Compounds in the Training Series and the Cross-Validation Series



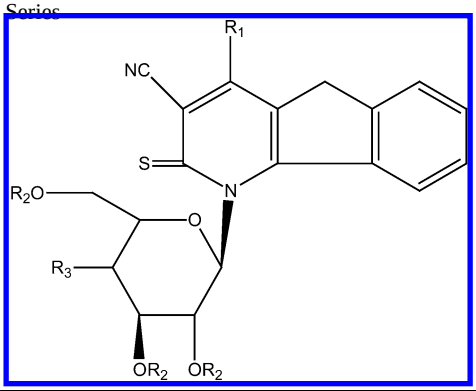
no.	R	S	T	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	act obs	act calc	prob
161[a]	S	C	N	N =	N	N =	C	CH =	C		=O		H	−[76]	−	7.57
162	O	C	N	CH =	C	N =	C	CH =	C		=O	4-NO_2Ph	H	−[76]	−	0.70
163[a]	S	C	N	CH =	C	N =	C	CH =	C		=O	4-NO_2Ph	H	−[76]	−	0.06
164	O	C	N	CH =	C	N =	C	CH =	C		=O	CH_2OH	H	−[76]	−	9.09
165[a]	S	C	N	CH =	C	N =	C	CH =	C		=O	CH_2OH	H	−[76]	−	1.82
166	CH	C =	C	CH =	C	O	C =	N	C	OH	H	$(CH_2)_9F$	=O	−[77]	−	4.52
167[a]	CH	C =	C	CH =	C	O	C =	N	C	OH	H	$(CH_2)_9Br$	=O	−[77]	−	4.67

[a] Compound in a cross-validation set.

lecular descriptors and the activity of the compounds under investigation. To obtain the QSAR model we calculate multiple 2D (topological) and 3D (topographic) indices as well as quantum chemical indices. Among the topological indices we calculate molecular connectivity indices[17,18] and Wiener[19] and Balaban[20] indices as well as edge connectivity indices[21−23] of different types and sizes. The topographic descriptors computed were only the so-called quantum-connectivity indices, which have been described and analyzed in detail in the recent literature.[24−28] Quantum-connectivity indices were calculated after the molecular structures of all compounds were optimized geometrically using semiempirical AM1 calculations using the HyperChem software.[29] The starting points for these calculations were the most stable conformations, as in previous structural and quantum chemical calculations.[30,31] The calculation of the different descriptors shown in Table 2 was performed with the system MODEST.[32]

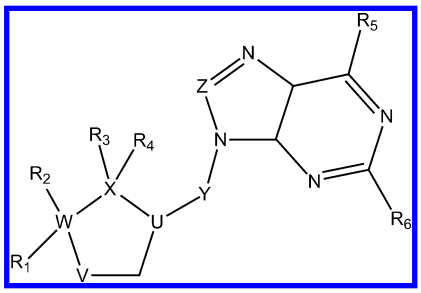
The QSAR models were selected from the best linear regression equations. These models were selected by fol-

Table 5. Molecular Structures and Anti-HIV Activities (Experimental and Calculated) for Selected Compounds in the Training Series



no.	R_1	R_2	R_3	act obs	act calc	prob
168	4-ClC_6H_4	$COCH_3$	αO−CO−CH_3	−[78]	+	64.81
169	4-$CH_3OC_6H_4$	$COCH_3$	αO−CO−CH_3	+[78]	+	73.21
170	4-ClC_6H_4	H	αOH	+[78]	−	23.60
171	4-$CH_3OC_6H_4$	H	αOH	−[78]	−	30.74
172	4-ClC_6H_4	$COCH_3$	βO−CO−CH_3	+[78]	+	59.74
173	4-$CH_3OC_6H_4$	$COCH_3$	βO−CO−CH_3	−[78]	+	68.00
174	4-ClC_6H_4	H	βOH	−[78]	−	19.70
175	4-$CH_3OC_6H_4$	H	βOH	−[78]	−	25.61

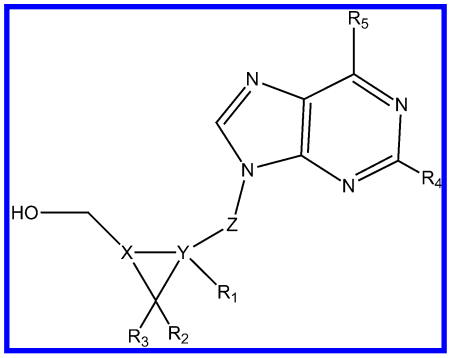
Table 6. Molecular Structures and Activities (Experimental and Calculated by TOPS-MODE) for Selected Compounds in the Training Series and External Prediction Series



no.	U	V	W	X	Y	Z	R_1	R_2	R_3	R_4	R_5	R_6	act obs	act calc	prob
176	CH(S)	O	C(R)	O	O	CH	H	CH_2OH			Cl	H	−[79]	−	5.73
177	CH(S)	S	C(S)	O	O	CH	H	CH_2OH			NH_2	H	−[79]	−	1.71
178	CH(R)	S	C(S)	O	O	CH	H	CH_2OH			OCH_3	H	−[79]	−	14.66
179	CH(R)	O	C(R)	O	O	CH	H	CH_2OH			OH	H	−[79]	+	67.44
180[a]	CH(S)	CH_2	C	C		N	H	H	CH_2OH	H	OH	H	−[30]	−	6.26
181[a]	CH(S)	CH_2	C	C	CH_2	N	H	H	CH_2OH	H	OH	H	−[30]	−	6.53
182[a]	CH(S)	CH_2	C	C		N	H	H	CH_2OH	H	NH_2	H	−[30]	−	6.17
183[a]	CH(S)	CH_2	C	C	CH_2	N	H	H	CH_2OH	H	NH_2	H	−[30]	−	7.01
184[a]	CH(S)	CH_2	C	C	CH_2	CH	H	H	CH_2OH	H	NH_2	H	−[30]	−	2.70
185[a]	CH(S)	CH_2	C	C	CH_2	CH	H	H	CH_2OH	H	OH	H	−[80]	−	3.04
186[a]	CH(S)	CH_2	C(R)	C	CH_2	CH	CH_3	CH_2OH	CH_3	CH_3	OH	NH_2	−[81]	−	0.27

[a] Compound in a external prediction set.

Table 7. Molecular Structures and Anti-HIV Activities for Selected Compounds in the Cross-Validation Set and External Prediction Set



no.	X	Y	Z	R_1	R_2	R_3	R_4	R_5	act obs	act calc	prob
187[a]	CH(R)	C=	CH		H	H	H	NH_2	−[65]	−	44.17
188[a]	CH(S)	C=	CH		H	H	H	NH_2	−[65]	−	36.94
189[b]	CH(R)	C(R)	CH_2	CH_2OH	H	H	H	NH_2	−[82]	−	8.01
190[b]	CH(R)	C=	CH		F	F	H	NH_2	−[83]	−	0.06
191[b]	CH(R)	C=	CH		F	F	NH_2	OH	−[83]	−	0.14
192[b]	CH(R)	C=	CH		F	F	NH_2	Cl	−[83]	−	0.02

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

lowing statistical criteria. We considered not only the quality of the regression fit measured by the correlation coefficient and the standard error but also the quality of the selected model in predicting the observed variable for external series of compounds. This predictive power of the model is measured by using a leave-one-out cross-validation experiment. We also carried out a randomization test for checking that our QSAR models are not produced by chance correlations. The STATISTICA package was also used for the development of the QSAR equation.

RESULTS AND DISCUSSION

Classification Function. The predictive equation derived from the STATISTICA package is as follows

$$\begin{aligned}\text{Class} = {} & 0.9945\mu_2 - 1.1708\mu_4 + 0.2806\mu_6 - \\ & 2.5598{*}10^{-2}\mu_8 + 8.9699{*}10^{-4}\mu_{10} - 2.0084{*}10^{-6}\mu_{13} + \\ & 7.5197{*}10^{-3}\mu_0\mu_2 - 8.1811{*}10^{-3}\mu_1\mu_2 + \\ & 8.3870{*}10^{-2}\mu_0 I_1 - 0.8605 I_2 + 0.1206 I_3 + 7.8048 \quad (2)\end{aligned}$$

$$N = 62\ \lambda = 0.68\ F(11,50) = 2.12\ D^2 = 2.08\ p < 0.05$$

where N is the number of compounds included in the discriminant analysis calculation, λ is the Wilks' statistics, F is the Fisher ratio and D^2 is the squared Mahalanobis distance.

The results obtained in the classification of the compounds that form the training series, the cross-validation series and the external prediction series are shown in Tables 3−8. When the differences in the classification percentage between activity and inactivity are below 5%, the compounds are considered as unclassified and indicated by the letter U in the tables. In this analysis there is also a series of compounds classified as false actives (i.e., compounds that are classified by the model as actives but are known to be inactives). Conversely, there is also a series of false inactives compounds. The percentage of unclassified compounds in the training set is only 3.2% (2/62), while the percentages of false actives and false inactives compounds are 9.68% (6/62) and 4.8% (3/62), respectively. The model classifies correctly 82.26% (51/62) of the compounds that form the training series. When the unclassified compounds are not taken into account the level of correct classification rises to 85% (51/60).

Chart 1

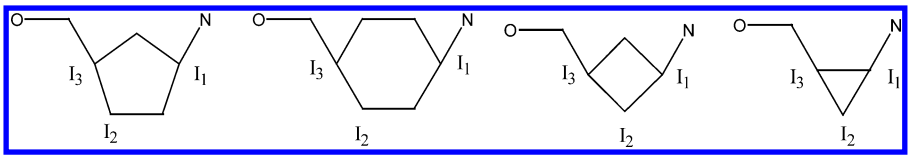
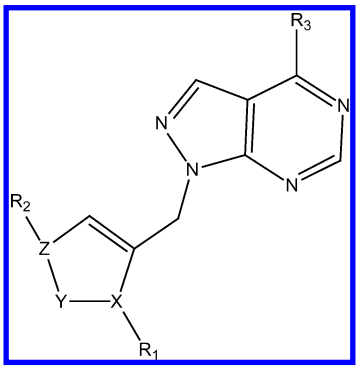


Table 8. Molecular Structures and Activities for Selected Compounds in the External Prediction Set



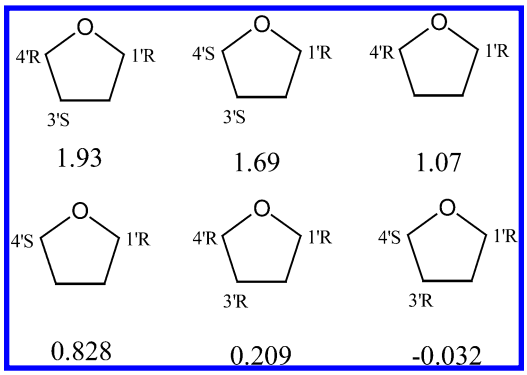
no.	X	Y	Z	R_1	R_2	R_3	act obs	act calc	prob
193	N =	N	N		$CH_2O(CH_2)_2OH$	SCH_3	–[84]	–	0.07
194	N =	N	N		$CH_2O(CH_2)_2OH$	$SCH_2C_6H_5$	–[84]	–	2.59
195	N =	N	N		$CH_2O(CH_2)_2OH$	NH_2	–[84]	–	7.51
196	N =	N	N		$CH_2O(CH_2)_2OH$	$NHCH_3$	–[84]	–	10.63
197	N =	N	N		$CH_2O(CH_2)_2OH$	$NHCH_2C_6H_5$	–[84]	–	23.21
198	N =	N	N		$CH_2O(CH_2)_2OH$	OCH_3	–[84]	–	8.18
199	N =	N	N		$CH_2O(CH_2)_2OH$	OH	–[84]	–	7.93
200	N	N =	N	$CH_2O(CH_2)_2OH$		SCH_3	–[84]	–	0.08
201	N	N =	N	$CH_2O(CH_2)_2OH$		$SCH_2C_6H_5$	–[84]	–	3.36
202	N	N =	N	$CH_2O(CH_2)_2OH$		NH_2	–[84]	–	8.13
203	N	N =	N	$CH_2O(CH_2)_2OH$		$NHCH_3$	–[84]	–	12.50
204	N	N =	N	$CH_2O(CH_2)_2OH$		$NHCH_2C_6H_5$	–[84]	–	24.35
205	N	N =	N	$CH_2O(CH_2)_2OH$		OCH_3	–[84]	–	8.66
206	N	N =	N	$CH_2O(CH_2)_2OH$		OH	–[84]	–	8.59

The proportion of compounds per variable included in this discriminant model is of 5.6. Due to this apparently large number of variables in the model we have made great emphasis in the study of the predictive power of this model. Consequently, we have first conducted a leave-one-out cross-validation test in which the classification of the compound left out is evaluated with a model related with the rest of compounds. This process was repeated as many times as compounds are in the training set. The percentage of good classification for this test does not differ significantly from the ones obtained for the original model (79% of global correct classification for the cross-validation versus 82% for the orginal model). A second validation test was then conducted using a cross-validation set of 35 compounds previously selected at random from the whole data set. The compounds used in this cross-validation set were never used in the training of the model developed. The percentages of false actives and false inactives compounds are 14.29% (5/35) and 2.86% (1/35), respectively. A total of 82.86% (29/35) of the compounds are correctly classified by the discriminant equation. At this point we also conducted a third and more definitive test for validating the predictive ability of this model. It consisted of selecting an external prediction set of nucleosides collected from the literature after the model was developed. This external prediction series comprised of 109 compounds (13 active and 96 inactive) and, of these, 85 were correctly classified. Thus, in the prediction series the percentage of good classification is 78%. The model correctly assigns 76% of the inactive compounds and 92.3% of the active compounds (see Tables 3−8). In closing, we have proved using three different approaches that this classification model is predictive and robust.

Concerning the number of variales in the model we would like to remark that "it was considered almost as a dogma by many investigators that a meaningful QSAR should be based on at least five observations per variable included. This is, of course, not correct. The ratio "number of observations per variable included" will simply be reflected in *F*- and *t*-statistics and in the confidence intervals of regression coefficients and the predicted values which increase as this ratio decreases".[33] In addition, it is well-known that when a cross-validation sample is specified, as we have done in the current work, best-subset selection can also be based on the misclassification rates for the cross-validation sample. In other words, after estimating the discriminant functions for a given set of predictors, the misclassification rates for the cross-validation sample are computed, and the model (subset of predictors) that yields the lowest misclassification rate for the cross-validation sample is chosen. This is a powerful technique for choosing models that may yield good predictive validity, while avoiding overfitting of the data.

Structural Interpretation. The Contribution of the Fragments. The TOPS-MODE approach has a number of advantages that allow for knowing the contributions of structural fragments in terms of the activity or biological

Chart 2



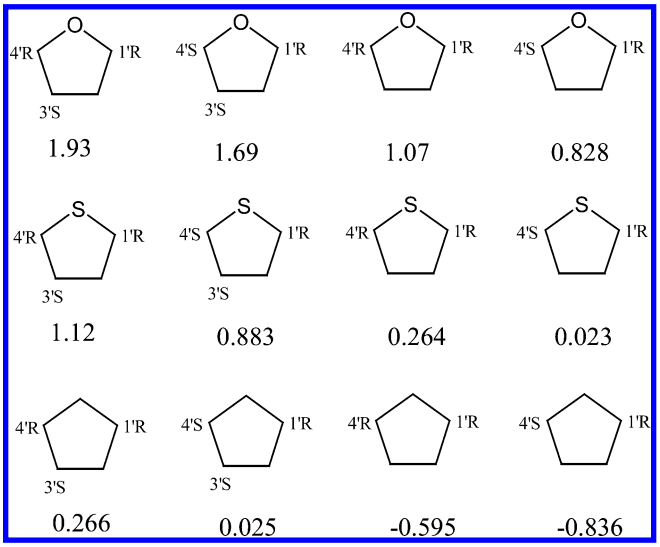
property under investigation.[13,14] The methodological aspects of this approach have been explained in detail in the literature.[85,86] In the study described here, the positive or negative contributions of various fragments is evaluated with respect to the activity using only the information apported by the spectral moments in the model. The contribution to the activity of a given fragment is considered in terms of the contribution of the substructure to the discriminant function. On the other hand, it is also of interest to identify the active fragments with the aim of discovering possible bidimensional pharmacophores that would allow the synthesis of novel nucleoside analogues with anti-HIV activity. In a similar way to the compounds involved in the analysis, the chirality of the stereocenters in the substructures was taken into account. In some cases significant differences were found in terms of the contribution of the stereocenters to the activity. The calculation of the different fragments allows a more detailed analysis of the importance that small changes at the substructural level have on the pharmacological action. Such a study allows several rules to be postulated that could direct the synthesis of novel nucleoside analogues:

(1) The preferred configurations of the sugar ring are 1′R, 3′S, 4′R and 1′R, 3′S, 4′S (see Chart 2). These are the configurations of active nucleosides such as FddGuo and FddDAPR (compounds **13** and **14** in Table 3). Typical nucleosides with an $-N_3$ group in the 3′-position, such as AzddGuo, AzddDAPR and AzddAdo (compounds **6**, **7** and **8**), show anti-HIV activity, and, in these cases, the configuration of the ring is 1′R, 3′S, 4′S.

(2) A change in the configuration at the 3′-position leads to a large change in the anti-HIV activity, with the compounds having a 3′S chiral center showing the highest activity (see Chart 3). Indeed, the change from a 3′S to a 3′R center gives rise to an inactive compound. When the $-\alpha N_3$ radical in the 3′-position in AzddAdo (compound **8**) is replaced by a $-\beta N_3$ group (compound **9**), the complete loss of anti-HIV activity results.

(3) The replacement of the ring oxygen by a sulfur atom or a methylene group causes a decrease in the activity of the compounds (see Chart 4). It can be seen from the results

Chart 4



in Tables 3−8 that all of the most active compounds have an oxygen in the ring. The dioxolane adenine nucleoside[45] has higher anti-HIV activity in MT-4 assays than (−)BCH-1230 (oxathiolane adenine nucleoside),[44] with values of $ED_{50} = 1.3$ μM and 4.34 μM, respectively.

(4) Five-membered rings contribute most to the activity and, along with systems containing four-membered rings, provide the most active compounds. On the other hand, nucleosides that contain six- or three-membered rings are absent from the compounds considered in the literature to be highly active (see Table 7). The contributions of the different fragments to the activity are shown in Chart 5.

(5) The contributions to anti-HIV activity of different five-membered rings was studied (see Chart 6). The highest contribution is given by dioxolane ring which, together with tetrahydrofuran derivatives, gives the most active nucleosides. The introduction of the sulfur atom in the 5′-position of the ring leads to the loss of anti-HIV activity in most of the purine-based nucleosides. A clear example of this effect is provided by compounds **33**, **3** and **95** (DXG, ddGuo and an oxathiolane derivative), which have activities of 0.05, 7.6 and >100 μM, respectively. It can be seen from structure **35** (Figure 1) how the sulfur atom contributes very little to the activity.

(6) Bearing in mind that the 3′-position of the furan ring is extremely important in terms of activity, the contributions of different substituents in this position were calculated. It can be seen from Chart 7 that comparable linear substituents favor activity to a greater extent than branched substituents, i.e., linear substituents are advantageous when comparing groups with similar electronic characteristics.

The TOPS-MODE approach also allows the contributions to the activity of different bonds to be calculated for a given

Chart 3

Chart 5

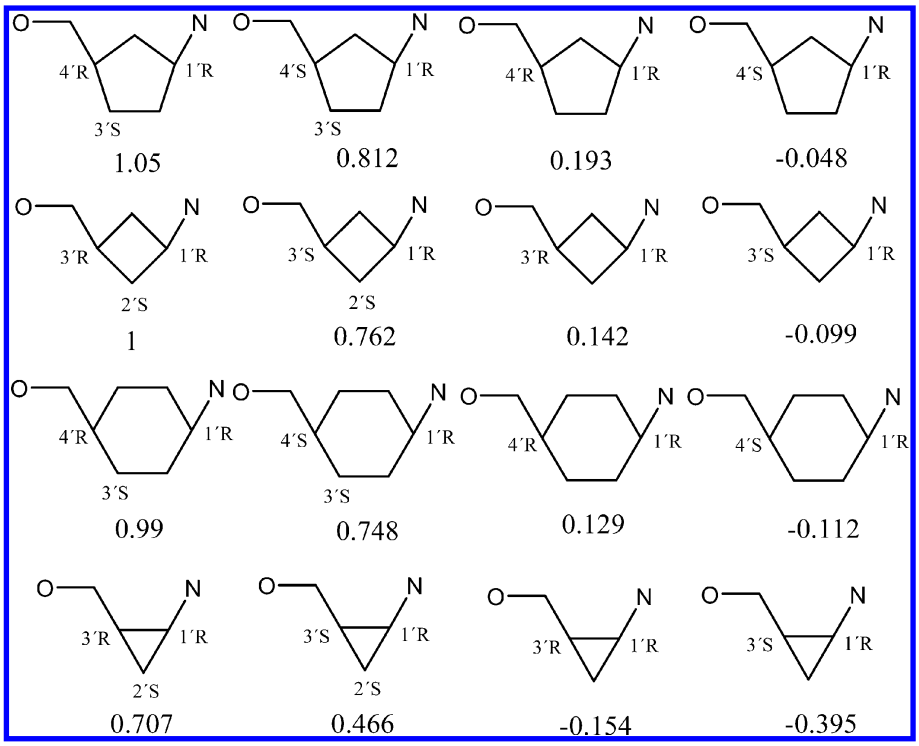
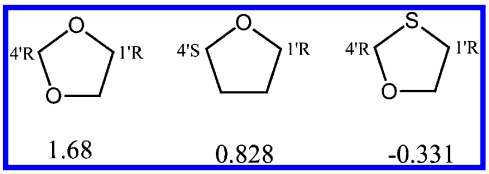


Chart 6



compound.[87,88] The total spectral moments can be expressed as a linear combination of the local or bond spectral moments. In this way it is possible to calculate the local spectral moments and substitute these for the total moments in eq 2. Five of the most interesting compounds are shown in Figure 1 along with the respective contributions of the bonds. It appears that the contribution of the sugar moiety is very significant, although the 2- and 6-positions of the purine unit also markedly influence the activity. The presence of an azide group in the 3′-position of the tetrahydrofuran ring makes a significant contribution to the anti-HIV activity. This effect is clearly shown by AzddDAPR and AzddGuo, with activities of 0.3 and 1.4 μM, respectively, while the nucleosides ddDAPR and ddGuo, which are not substituted in the 3′-position, give lower activities of 3.6 and 7.6 μM, respectively. Another substituent of potential interest in the 3′-position is fluorine. Despite the fact that it was found to be of great interest in pyrimidine nucleosides, our calculations show that the 3′-fluoro substituent has a negative contribution to anti-HIV activity. This finding is supported by the data for compounds **1** and **2** (ddAdo and ddIno), whereas FddAdo and FddIno (compounds **11** and **17**) show activities of >10 μM. Another important aspect concerns the purine base and sugar bond. In this sense, lengthening of the chain gives rise to a significant decrease in activity (see Tables 6−8).

The contribution to the activity of different substituents allows the proposal of various hypotheses aimed at rationalizing the design of this type of pharmacological agent. In this respect the TOPS-MODE approach has a significant role to play in the development of compounds with anti-HIV activity.[89,90]

QSAR Model. In an effort to obtain a quantitative model for the prediction of anti-HIV activity in MT-4 assays, a QSAR equation was developed from the previously calculated descriptors (Table 2). The activity values observed (Log $1/EC_{50}$) are shown in Table 9 along with those calculated using the model as well as residual and delete residual. It can be seen from this table that a series of compounds from the chemical literature has been introduced into the model and that these compounds have similar structural characteristics. Leave-one-out cross-validation was used in the analysis to corroborate the results obtained with the QSAR function. Three 3D connectivity indices, two 2D connectivity indices and a quantum chemical index were used in the QSAR equation. The linear regression equation and the statistical parameters are as follows:

$$\text{Log } 1/EC_{50} = -0.231(\pm 0.605) + 1.242(\pm 0.276)E_{LUMO} - 6.775(\pm 0.583)[{}^4\Omega_p] - 4.172(\pm 0.709)[{}^6\Omega_p] + 13.102(\pm 1.148)[{}^5\epsilon_p(\rho)] - 10.609(\pm 0.760)[{}^5\epsilon_p] + 7.927(\pm 0.581)[{}^4\chi_p]$$

$$n = 30 \; R = 0.958 \; s = 0.199 \; RMSECV = 0.239$$

In this equation R is the correlation coefficient, s is the standard deviation, n is the number of compounds included in the model and *RMSECV* is the mean squared error of the leave-one-out cross-validation. The results of a randomization test shown in Figure 2 demonstrate that this QSAR model is not a chance correlation showing that the correlation coefficient of the regression and of the cross-validation are very far away from those obtained for models with random variables.

The correlation between the different variables introduced into the model is high giving the difficult interpretation of

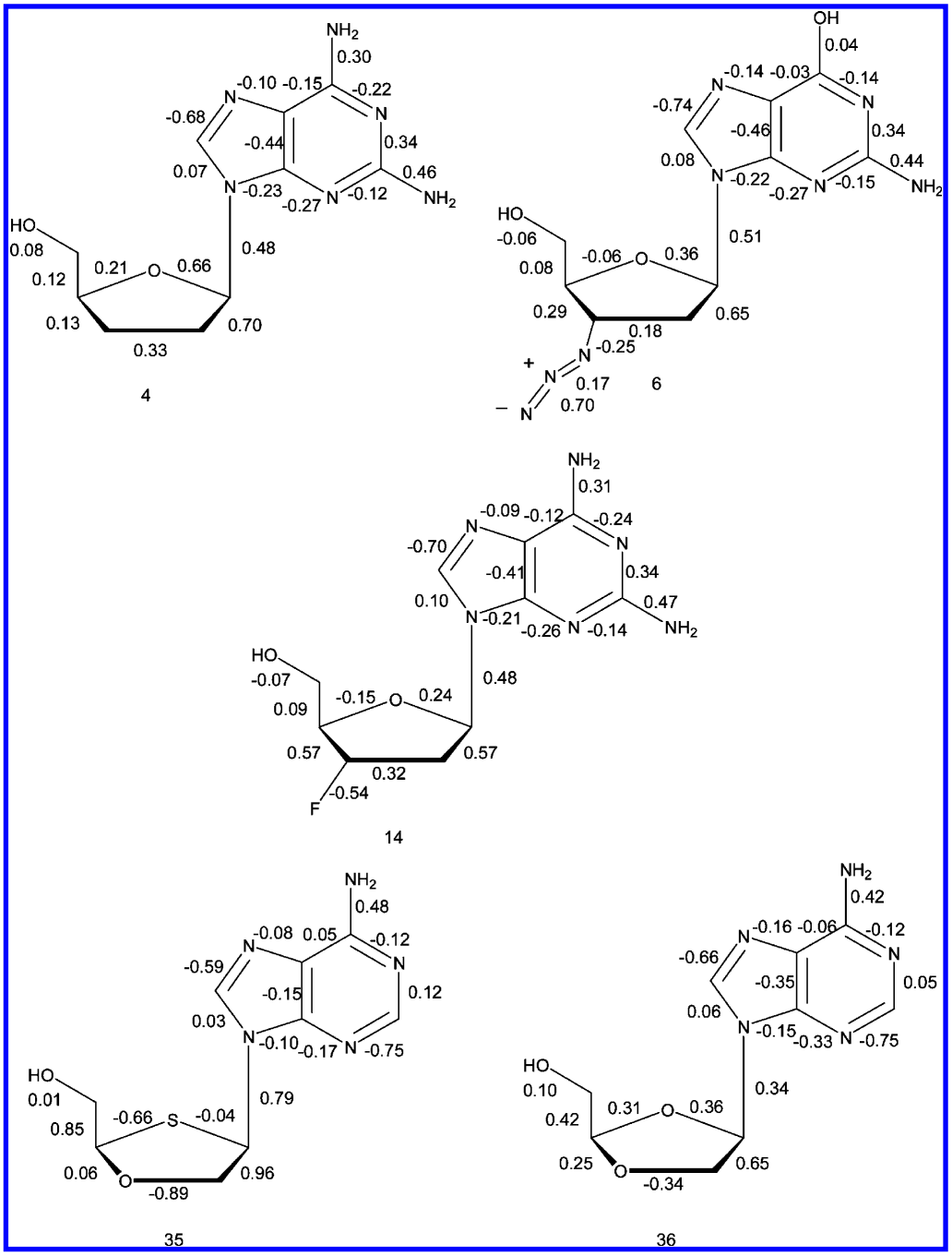
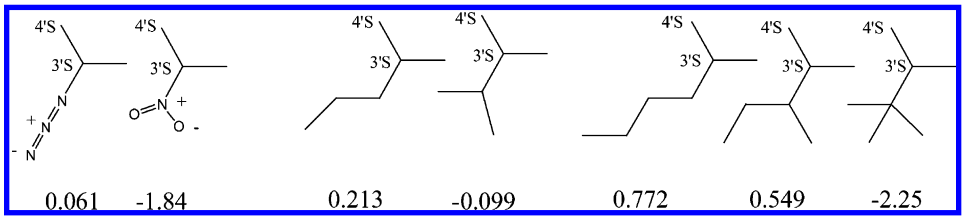


Figure 1. Bond contributions to the anti-HIV activity obtained from the classification model (2) for selected compounds in the database.

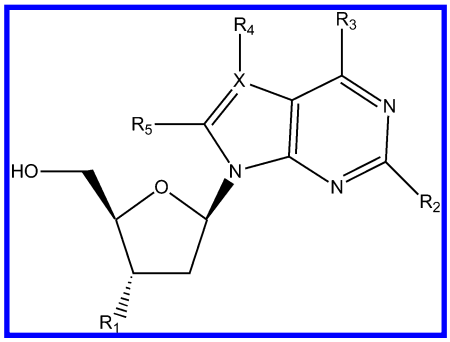
Chart 7



the QSAR and the instability of the regression coefficients, for which it was necessary to employ a variable orthogonalization method. The information in Tables10 and 11 shows the correlation between the molecular descriptors used in the QSAR equation and the regression coefficients of the orthogonalized descriptors obtained using the method described by Randić.[91,92] The importance of this model lies in its capacity for the quantitative prediction of anti-HIV activity in purine based nucleosides in MT-4 assays.

A mechanistic interpretation of this QSAR is not excepted of risk. There are several factors contributing to this difficulty in the interpretation. The most important one is related to the fact that the biological test used to generate the EC_{50} values does not account only for the drug-receptor interac-

Table 9. Experimental and Predicted Values of the log $1/EC_{50}$ in the MT-4 Assay for Compounds Used in the QSAR Model as Well as Residual and Leaves-One-Out Cross-Validation (Delete Residual)



no. (ref)	R_1	R_2	R_3	R_4	R_5	X	log $1/EC_{50}$ mol/L obsd value	log $1/EC_{50}$ mol/L pred value	residual	deleted residual
1′[34]	H	NH_2	OH	-	H	N	5.12	5.16	−0.04	−0.05
2′[34]	H	NH_2	NH_2	-	H	N	5.44	5.59	−0.15	−0.23
3′[35,62]	H	OH	OH	-	H	N	5.00	4.77	0.23	0.28
4′[36]	H	H	NH_2	-	Br	N	3.32	3.32	−0.00	−0.01
5′[34]	N_3	NH_2	OH	-	H	N	5.85	5.93	−0.08	−0.09
6′[34]	N_3	NH_2	NH_2	-	H	N	6.52	6.23	0.29	0.39
7′[36]	N_3	H	NH_2	-	H	N	5.30	5.23	0.07	0.09
8′[34]	F	H	NH_2	-	H	N	4.30	4.27	0.03	0.04
9′[34]	F	NH_2	NH_2	-	H	N	5.35	5.41	−0.06	−0.09
10′[36]	F	H	OH	-	H	N	3.65	3.76	−0.11	−0.17
11′[48]	H	H	NH_2	CN	H	C	5.07	4.86	0.21	0.28
12′[48]	H	H	NH_2	$CONH_2$	H	C	4.19	4.11	0.08	0.12
13′[66]	N_3	NH_2	OMe	-	H	N	5.05	5.14	−0.09	−0.10
14′[66]	N_3	NH_2	OC_2H_5	-	H	N	4.64	4.85	−0.21	−0.22
15′[66]	N_3	NH_2	OC_3H_7	-	H	N	5.05	4.76	0.29	0.34
16′[66]	N_3	NH_2	$OCHMe_2$	-	H	N	4.66	4.70	−0.04	−0.06
17′[66]	N_3	NH_2	OC_4H_9	-	H	N	4.72	4.74	−0.02	−0.03
18′[66]	N_3	NH_2	OC_6H_5	-	H	N	4.37	4.11	0.26	0.41
19′[66]	N_3	NH_2	NHC_2H_5	-	H	N	4.77	4.94	−0.17	−0.18
20′[66]	N_3	NH_2	NHC_3H_7	-	H	N	4.64	4.79	−0.15	−0.17
21′[66]	N_3	NH_2	NHC_4H_9	-	H	N	4.54	4.80	−0.26	−0.29
22′[66]	N_3	NH_2	NHC_6H_5	-	H	N	3.95	4.24	−0.29	−0.46
23′[66]	N_3	NH_2	$NH(CH_2)_2$-C_6H_5	-	H	N	4.68	4.70	−0.02	−0.03
24′[66]	N_3	NH_2	N(Me)-C_2H_5	-	H	N	4.96	5.17	−0.21	−0.22
25′[66]	N_3	NH_2	NC_3H_6	-	H	N	5.05	4.81	0.24	0.28
26′[66]	N_3	NH_2	NC_4H_8	-	H	N	5.10	5.03	0.07	0.11
27′[66]	N_3	NH_2	Cl	-	H	N	5.22	5.46	−0.24	−0.47
28′[67]	N_3	NH_2	NH-*c*-Pr	-	H	N	4.74	4.81	−0.07	−0.07
29′[67]	N_3	NH_2	NH-*c*-Bu	-	H	N	4.85	4.66	0.19	0.24
30′[67]	N_3	NH_2	N(Me)-*c*-Pr	-	H	N	5.40	5.17	0.23	0.30
31′[34]	H	H	NH_2[a]	-	H	N	5.19	4.46	0.73	
32′[34]	H	H	OH[a]	-	H	N	5.00	3.96	1.04	
33′[34]	F	NH_2	OH[a]	-	H	N	5.62	4.99	0.63	
34′[66]	N_3	NH_2	OCH_2-C_6H_5[a]	-	H	N	5.10	4.33	0.77	
35′[66]	N_3	NH_2	NHMe[a]	-	H	N	4.28	5.12	−0.84	
36′[66]	N_3	NH_2	NMe_2[a]	-	H	N	4.92	5.42	−0.50	

[a] Compounds that are considered outliers (standardized residual method) in the QSAR model.

tion, i.e., the interaction of the nucleoside with the HIV reverse-transcriptase. In the MT-4 assay the values of biological response depend not only on this interaction but also on the capacity of the chemical to penetrate the cell, to distribute inside it as well as of being metabolically activated by the phosphorylation reactions at the hydroxymethyl end. Consequently, to analyze whether some variables are influencing one or another process is very speculative. Thus, we consider the usability of this QSAR model on the most pragmatic side of the problem using it as a tool for selecting and/or designing potent anti-HIV compounds in which all possible factors influencing the biological response, i.e., partition, distribution, metabolism, interaction, etc., are taken into account.

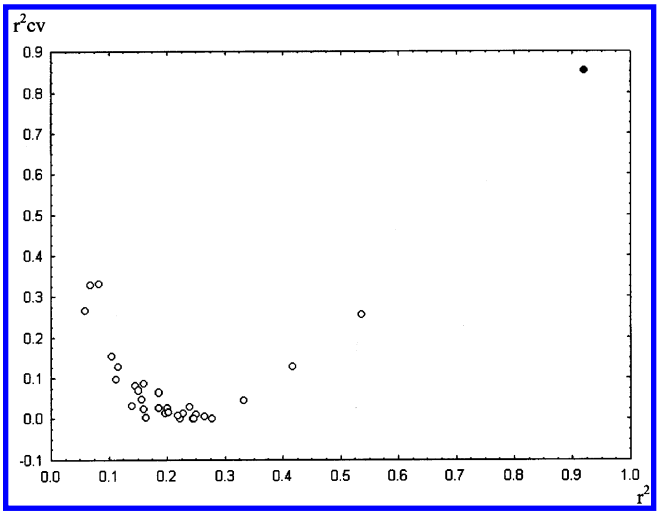


Figure 2. Randomization test in order to test possible chance correlations. Correlation coefficient of the regression and of the cross-validation are very far away from those obtained for models with random variables (in black QSAR model).

Table 10. Correlation Coefficients of the Six Variables (Molecular Descriptors) in the QSAR Model

	E_{LUMO}	$^4\Omega_p$	$^6\Omega_p$	$^5\epsilon_p(\rho)$	$^5\epsilon_p$	$^4\chi_p$
E_{LUMO}	1.00	0.33	0.36	0.29	0.22	0.27
$^4\Omega_p$		1.00	0.98	0.97	0.92	0.96
$^6\Omega_p$			1.00	0.97	0.91	0.94
$^5\epsilon_p(\rho)$				1.00	0.98	0.98
$^5\epsilon_p$					1.00	0.98
$^4\chi_p$						1.00

Table 11. Regression Coefficients in the QSAR Model for Orthogonal Molecular Descriptors

E_{LUMO}	$5\epsilon_p$	$4\chi_p$	$4\Omega_p$	$5\epsilon_p(\rho)$	$6\Omega_p$	intercept
1.6436						5.1673
1.6436	−0.1640					5.1673
1.6436	−0.1640	1.4111				5.1673
1.6436	−0.1640	1.4111	−0.8677			5.1673
1.6436	−0.1640	1.4111	−0.8677	8.0125		5.1673
1.6436	−0.1640	1.4111	−0.8677	8.0125	−4.1724	5.1673

CONCLUSION

The graph theoretical approach and the use of different indices to describe compounds, e.g., spectral moments, has greatly helped in the understanding and design of novel compounds that show biological activity. Once again, TOPS-MODE has allowed the design of a predictive model for anti-HIV activity in nucleoside analogues. This model enables us to evaluate the biological potential of a novel nucleoside analogue prior to synthesis. In this way it is possible to develop a rational synthesis program that excludes compounds that have a low probability of being active against the HIV. Furthermore, the use of QSAR enables the biological activity of a novel compound to be optimized through the selection of the most appropriate substituents and provides quantitative data concerning activity.

One of the advantages of TOPS-MODE is the possibility of calculating the contribution to the activity (or other biological property) of a range of fragments or molecular substructures. This ability represents a powerful tool in the

design of novel pharmacological agents in that it provides information concerning the most suitable modifications in various positions of the nucleosides.

REFERENCES AND NOTES

(1) Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem., Int. Ed. Engl*. **1990**, *29*, 992−1023.

(2) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem*. **1983**, *4*, 187−217.

(3) Segal, G. A. *Methods of electronic structure calculations*; Plenum Press: New York, 1972.

(4) 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.

(5) Estrada, E.; Peña, A. In Silico studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem*. **2000**, *8*, 2755−2770.

(6) 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.

(7) Montagnier, L.; Dauguet, C.; Axler, C.; Chamaret, S.; Gruest, J.; Nugeyre, M. T.; Rey, F.; Barré-Sinoussi, F.; Chermann, J. C. A new type of retrovirus isolated from patients presenting with lymphadenopathy and AIDS: structural and antigenic relatedness with equine infectious anemia virus. *Ann. Inst. Pasteur. Virol*. **1984**, *135E(nE1)*, 119−134.

(8) Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B.; White, G.; Foster, P.; Markham, P. D. Frequent detection and isolation of cytopathic retroviruses (HTLV−III) from patients with AIDS and at risk for AIDS. *Science* **1984**, *224*, 500−503.

(9) Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; St Clair, M. H.; Nusinoff-Lehrman, S.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S. 3′-Azido-3′-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7096−7100.

(10) Miysuya, H.; Broder, S. Inhibition of the in vitro infectivity and cytopathic effect of human T-lymphotrophic virus type III/lymphadenopathy-associated virus (HTLV−III/LAV) by 2′, 3′-dideoxynucleosides. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 1911−1915.

(11) Schinazi, R. F. Competitive inhibitors of human inmunodeficiency virus reverse transcriptase. *Perspect. Drug Discovery Des*. **1993**, *1*, 151−180.

(12) Baba, M.; Pauwels, R.; Balzarini, J.; Herdewijn, P.; De Clercq, E.Selective inhibition of human immunodeficiency virus (HIV) by 3′-azido-2′, 3′-dideoxyguanosine in vitro. *Biochem. Biophys. Res. Commun*. **1987**, *145*, 1080−1086.

(13) 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.

(14) 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.

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

(16) STATISTICA 6.0; StatSoft Inc.: Tulsa, U.S.A., 2002.

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

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

(19) Wiener, H. Structural determination of paraffins boilings points. *J. Am. Chem. Soc*. **1947**, *69*, 17−20.

(20) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett*. **1982**, *89*, 399−404.

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

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

(23) 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.

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

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

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

(27) Estrada, E.; Molina, E. 3D connectivity indices in QSPR/QSAR studies. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 791−797.

(28) Estrada, E.; Delgado, E. J.; Alderete, J. B.; Jaña, G. A. Quantum-connectivity descriptors in modeling solubility of environmentally important organic compounds. *J. Comput. Chem*. **2004**, *25*, 1787−1796.

(29) HyperChem 3.0. Hypercube, Inc and Autodesk, Inc. 1993.

(30) Santana, L.; Teijeira, M.; Uriarte, E.; Balzarini, J.; De Clercq, E. Synthesis, conformational analysis and antiviral and antitumoral activity of new 1,2-disubstituted carbocyclic nucleosides. *Eur. J. Med. Chem*. **2002**, *37*, 755−760.

(31) Teran, C.; Santana, L.; Teijeira, M.; Uriarte, E.; De Clercq, E. Design, synthesis, conformational analysis and biological activities of purine-based 1,2-di-substituted carbocyclic nucleosides. *Chem. Pharm. Bull*. **2000**, *48*, 293−295.

(32) Rodríguez, L.; Estrada, E. MODEST (MOlecular DESign Tool) for Windows. Version 3.0. 2000.

(33) Franke, R. *Theoretical drug design methods*; Elsevier: Amsterdam, 1984; p 167.

(34) De Clercq, E. HIV inhibitors targeted at the reverse transcriptase. *AIDS Res. Hum. Retrov*. **1992**, *8*, 119−133

(35) De Clercq, E. *Design of anti-aids drugs*; Pharmacochem. Libr.: 1990; Vol. 14, pp 141−194.

(36) 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.

(37) Herdewijn, P.; Pauwels, R.; Baba, M.; Balzarini, J.; De Clercq, E. Synthesis and anti-HIV activity of various 2′- and 3′-substituted 2′, 3′-dideoxyadenosines: a structure−activity analysis. *J. Med. Chem*. **1987**, *30*, 2131−2137.

(38) Herdewijn, P.; Balzarini, J.; Baba, M.; Pauwels, R.; Aerschot, A. V.; 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.

(39) Herdewijn, P.; Balzarini, J.; Pauwels, R.; Janssen, G.; Van Aerschot, A.; De Clercq, E. Synthesis and biological activity of the mono- and diamino analogues of 2′-deoxyadenosine, cordycepin, 9-(3-deoxy-D-threo-pentofuranosyl)adenine (a structural component of agrocin 84) and 9-(2-deoxy-D-threo-pentofuranosyl)adenine. *Nucleosides Nucleotides* **1989**, *8*, 1231−1257.

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

(41) Jeong, L. S.; 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.

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

(43) Gu, Z.; Wainberg, M. A.; Nguyen-Ba, P.; L'Heureux, L.; de Muys, J.-M.; Rando, R. F. Anti-HIV-1 activities of 1,3-dioxolane guanine and 2,6-diaminopurine dioxolane. *Nucleosides Nucleotides* **1999**, *18*, 891−892.

(44) 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.

(45) Siddiqui, M. A.; Brown, W. L.; Nguyen-Ba, N.; Dixit, D. M.; Mansour, T. S. Antiviral optically pure dioxolane purine nucleoside analogues. *Bioorg. Med. Chem. Lett*. **1993**, *3*, 1543−1546.

(46) Daluge, S. M.; Good, S. S.; Faletto, M. B.; Miller, W. H.; Clair, M. H. ST.; Boone, L. R.; Tisdale, M.; Parry, N. R.; Reardon, J. E.; Dornsife, R. E.; Averett, D. R.; Krenitsky, T. A. 1592U89, a viral carbocyclic nucleoside analogue with potent, selective anti-human immunodeficiency virus activity. *Antimicrob. Agents Chemother*. **1997**, *41*, 1082−1093.

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

(48) Pauwels, R.; Baba, M.; Balzarini, J.; Herdewijn, P.; Desmyter, J.; Robins, M. J.; Zou, R.; Madej, D.; De Clercq, E. Investigation on the anti-HIV activity of 2′,3′-dideoxyadenosine analogues with modifications in either the pentose or purine moiety. *Biochem. Pharmacol*. **1988**, *37*, 1317−1325.

(49) Katagiri, N.; Nomura, M.; Sato, H.; Kaneko, C.; Yusa, K.; Tsuruo, T.Synthesis and anti-HIV activity of 9-[*c*-4,*t*-5-bis(hydroxymethyl)-cyclopent-2-en-*r*-1-yl]-9*H*-adenine. *J. Med. Chem*. **1992**, *35*, 1882−1886.

(50) Rosenquist, A.; Kvarnström, I.; Classon, B.; Samuelsson, B. Synthesis of enantiomerically pure Bis(hydroxymethyl)-branched cyclohexenyl and cyclohexyl purines as potential inhibitors of HIV. *J. Org. Chem.* **1996**, *61*, 6282−6288.

(51) Shuto, S.; Obara, T.; Toriya, M.; Hosoya, M.; Snoeck, R.; Andrei, G.; Balzarini, J.; De Clercq, E.New neplanocin analogues. 1. Synthesis of 6′-modified neplanocin A derivatives as broad-spectrum antiviral agents. *J. Med. Chem.* **1992**, *35*, 324−331.

(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) Molas, M. P.; Matheu, M. I.; Castillón, S.; Isac-García, J.; Hernández-Mateo, F.; Calvo-Flores, F. G.; Santoyo-González, 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.

(54) Oohashi, T.; Nishiyama, S.; Yamamura, S.; Kato, K. Synthesis of [(2′S, 3′S)-bis(hydroxylmethyl)pyrrolidin-1-yl] purine and pyrimidine nucleosides as potential antiviral agents. *Bioorg. Med. Chem. Lett*. **1998**, *8*, 1187−1188.

(55) Wachtmeister, J.; Classon, B.; Samuelsson, B.; Kvarnström, I. Synthesis of carbocyclic 2′,3′-dideoxy-2′-fluoro-3′-C-hydroxymethyl nucleoside analogues as potential inhibitors of HIV and HSV. *Tetrahedron* **1997**, *53*, 1861−1872.

(56) Secrist III, J. A.; Riggs, R. M.; Tiwari, K. N.; Montgomery, J. A. Synthesis and anti-HIV activity of 4′-thio-2′,3′-dideoxynucleosides. *J. Med. Chem.* **1992**, *35*, 533−538.

(57) Kotra, L. P.; Newton, M. G.; Chu, C. K. Synthesis of 2,3-dideoxy-2,2-difluoro-L-*glycero*-pentofuranosyl nucleosides. *Carbohydr. Res*. **1998**, *306*, 69−80.

(58) Maguire, A. R.; Meng, W.; Roberts, S. M.; Willetts, A. J. Synthetic approaches towards nucleocidin and selected analogues; anti-HIV activity in 4′-fluorinated nucleoside derivates. *J. Chem. Soc., Perkin Trans 1* **1993**, *15*, 1795−1808.

(59) Boumchita, H.; Legraverend, M.; Zerial, A.; Lemaitre, M.; Huel, C.; Bisagni, E. New cyclobutyl analogues of nucleosides. Part 2. Synthesis and antiviral evaluation of 2-amino-7-[3,3-bis(hydroxymethyl)cyclobut-1-yl]-3*H*, 7*H*-pyrrolo[2,3-*d*]pyrimidin-4-one and of cyclobutyl analogues of the pyrimidine nucleosides. *Eur. J. Med. Chem*. **1991**, *26*, 613−617.

(60) Gourdel-Martin, M. E.; Huet, F. Synthesis of *cis*-disubstituted cyclobutenyl nucleoside analogues. *Nucleosides Nucleotides* **1999**, *18*, 645−648.

(61) Moon, H. R.; Kim, H. O.; Chun, M. W.; Jeong, L. S.; Marquez, V. E. Synthesis of cyclopropyl-fused carbocyclic nucleosides via the regioselective opening of cyclic sulfites. *J. Org. Chem*. **1999**, *64*, 4733−4741.

(62) Dimoglo, A. S.; Gorbachov, M. Y.; Lesnik, T. I.; Saraçoglu, M.; Gûzel, Y.; Yildirim, 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.

(63) Jeong, L. S.; Moon, H. R.; Yoo, S. J.; Lee, S. N.; Kim, H.-D.; Chun, M. W. Synthesis and antiviral activity of L-2′-deoxy-2′-up-fluoro-4′-thionucleosides. *Nucleosides Nucleotides* **1999**, *18*, 571−572.

(64) Kim, W.; Kim, H.; Rhee, H. An efficient synthesis of (±)-*cis*-2-amino-6-hydroxy-9-[4′-hydroxyethyl-2′-cyclopenten-1′-yl] purine. *Heterocycles* **2000**, *53*, 219−224.

(65) Cheng, C.; Shimo, T.; Somekawa, K.; Baba, M. 9-Hydroxymethylcyclopropylidenemethylenyladenine: the design, facile synthesis, isomer separation and anti-HIV-1 activities. *Tetrahedron* **1998**, *54*, 2031−2040.

(66) Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K.; Hansch, C. Comparative quantitative structure−activity relatioship studies on anti-HIV drugs. *Chem. Rev*. **1999**, *99*, 3525−3601.

(67) Freeman, G. A.; Shaver, S. R.; Rideout, J. L.; Short, S. A. 2-Amino-9-(3-azido-2,3-dideoxy-β-D-*erythro*-pentofuranosyl)-6-substituted-9H-purines: Synthesis and anti-HIV activity. *Bioorg. Med. Chem*. **1995**, *3*, 447−458.

(68) Gudmundsson, K. S.; Daluge, S. M.; Johnson, L. C.; Jansen, R.; Hazen, R.; Condreay, L. D.; McGuigan, C. Phosphoramidate protides of 2',3'-dideoxy-3'-fluoroadenosine and related nucleosides with potent activity against HIV and HBV. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 1953−1961.

(69) Gudmundsson, K. S.; Daluge, S. M.; Condreay, L. D.; Jonhson, L. C. Synthesis of novel 8-substituted carbocyclic analogues of 2',3'-dideoxyadenosine with activity against hepatitis B virus. *Nucleosides, Nucleotides Nucleic Acids* **2002**, *21*, 891−901.

(70) Franzyk, H.; Jensen, S. R.; Olsen, C. E.; Rasmussen, J. H. Synthesis of novel hydroxymethyl substituted analogues related to carbovir and neplanocin A. *Nucleosides, Nucleotides Nucleic Acids* **2002**, *21*, 23−43.

(71) Takamatsu, S.; Izawa, K.; Maruyama, T.; Katayama, S.; Hirose, N.; De Clercq, E. Synthesis and in vitro antiviral activity evaluation of 9-(2-azido-2,3-dideoxy-D-*threo*-pentofuranosyl)adenine derivatives. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 1053−1057.

(72) Gudmundsson, K. S.; Wang, Z.; Daluge, S. M.; Feldman, P. L. Synthesis of carbocyclic analogues of 2,3-dideoxysangivamycin, 2,3-dideoxytoyocamycin, and 2,3-dideoxytriciribine. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 1823−1830.

(73) Choi, Y.; Sun, G.; George, C.; Nicklaus, M. C.; Kelley, J. A.; Marquez, V. E. Synthesis and conformational analysis of a locked analogue of carbovir built on a bicyclo[3.1.0]hex-2-enyl template. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 2077−2091.

(74) Jeannot, F.; Gosselin, G.; Mathé, C. Synthesis and studies of 3'-C-Trifluoromethyl-β-D-ribonucleosides bearing the five naturally ocurring nucleic acid bases. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 2195−2202.

(75) Nair, V.; Mickle, T.; Bera, S. Cyclopropyl and related analogues of the anti-HIV compound, isodideoxyadenosine. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 239−247.

(76) 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.

(77) Brancale, A.; McGuigan, C.; Andrei, G.; Snoeck, R.; De Clercq, E.; Balzarini, J. Bicyclic nucleoside inhibitors of Varicella-Zoster virus (VZV): the effect of a terminal halogen substitution in the side-chain. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1215−1217.

(78) Elgemeie, G. E. H.; Mansour, O. A.; Metwally, N. H. Synthesis and anti-HIV activity of different novel nonclassical nucleosides. *Nucleosides Nucleotides* **1999**, *18*, 113−123.

(79) Nguyen-Ba, N.; Lee, N.; Chan, L.; Zacharie, B. Synthesis and antiviral activities of N-9-oxypurine 1,3-dioxolane and 1,3-oxathiolane nucleosides. *Bioorg. Med. Chem. Lett*. **2000**, *10*, 2223−2226.

(80) Santana, L.; Teijeira, M.; Uriarte, E.; Terán, C.; Andrei, G.; Snoeck, R.; Balzarini, J.; De Clercq, E. Synthesis and biological evaluation of 1,2-disubstituted carbonucleosides of 6-substituted purine and 8-azapurine. *Nucleosides Nucleotides* **1999**, *18*, 733−734.

(81) Blanco, J. M.; Caamaño, O.; Fernández, F.; Gómez, G.; Nieto, M. I.; Balzarini, J.; Padalko, E.; De Clercq, E. Synthesis and antiviral and cytostatic activities of carbocyclic nucleosides incorporating a modified cyclopentane ring. 1: guanosine analogues. *Nucleosides Nucleotides* **1997**, *16*, 159−171.

(82) Sekiyama, T.; Hatsuya, S.; Tanaka, Y.; Uchiyama, M.; Ono, N.; Iwayama, S.; Oikawa, M.; Suzuki, K.; Okunishi, M.; Tsuji, T. Synthesis and antiviral activity of novel acyclic nucleosides: Discovery of a cyclopropyl nucleoside with potent inhibitory activity against Herpes viruses. *J. Med. Chem*. **1998**, *41*, 1284−1298.

(83) Wang, R.; Ksebati, M. B.; Drach, J. C.; Zemlicka, J. Synthesis and antiviral activity of methylenedifluorocyclopropane analogues of nucleosides. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 329−332.

(84) Moukha-Chafiq, O.; Taha, M. L.; Lazrek, H. B.; Pannecouque, C.; Witvrouw, M.; De Clercq, E.; Barascut, J. L.; Imbach, J. L. Synthesis and biological activity of 4-substituted 1-[1-(2-hydroxyethoxy)-methyl-1,2,3-triazol-(4, 5)-ylmethyl]-1*H*-pyrazolo[3, 4-*d*]pyrimidines. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 1797−1810.

(85) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem*. **2001**, *8*, 1573−1588.

(86) Estrada, E. Extracting structural information from molecular graphs. The TOPS-MODE approach. *Adv. Quantum Chem*. **2005**, in press.

(87) Estrada, E.; Uriarte, E.; Gutierrez, Y.; Gonzalez, H. Quantitative structure-toxicity relationships using TOPS-MODE. 3. Structural factors influencing the permeability of commercial solvents through living human skin. *SAR QSAR Environ. Res*. **2003**, *14*, 145−163.

(88) Estrada, E.; Quincoces, J. A.; Patlewicz, G. Creating molecular diversity from antioxidants in Brazilian propolis. Combination of TOPS-MODE QSAR and virtual structure generation. *Mol. Diversity* **2004**, *8*, 21−33.

(89) Estrada, E.; Vilar, S.; Uriarte, E.; Gutierrez, Y. 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. *J. Chem. Inf. Comput. Sci*. **2002**, *42*, 1194−1203.

(90) Kodama, E.; Ohrui, H.; Sakata, S.; Kohgo, S.; Matsuoka, M.; Mitsuya, H.; Yamada, K. *4′-C-cyano-2′-deoxypurine nucleosides*. Patent WO 03068796, 2003.

(91) Randić, M. Orthogonal molecular descriptors. *New J. Chem*. **1991**, *15*, 517−525.

(92) Randić, M. Fitting of nonlinear regressions by orthogonalized power series. *J. Comput. Chem*. **1993**, *14*, 363−370.

CI049662O