A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds

Ernesto Estrada,*,† Eugenio Uriarte,† Alina Montero,‡ Marta Teijeira,† Lourdes Santana,† and Erik De Clercq§

Faculty of Pharmacy, Department of Organic Chemistry, Universidad de Santiago de Compostela, 15706 Santiago de Compostela, Spain, Centro de Bioactivos Químicos, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba, and Rega Institute for Medical Research, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

Received September 29, 1999

A topological substructural approach to molecular design (TOSS-MODE) has been introduced for the selection and design of anticancer compounds. A quantitative model that discriminates anticancer compounds from the inactive ones in a training series was obtained. This model permits the correct classification of 91.43% of compounds in an external prediction set with only 1.43% of false actives and 7.14% of false inactives. The model developed is then used in a simulation of a virtual search for Ras FTase inhibitors; 87% of the Ras FTase inhibitors used in this simulated search were correctly classified, thus indicating the ability of the TOSS-MODE model of finding lead compounds with novel structures and mechanism of action. Finally, a series of carbonucleosides was designed, and the compounds were classified as active/inactive anticancer compounds by using the model developed here. From the compounds so-designed, 20 were synthesized and evaluated experimentally for their antitumor effects on the proliferation of murine leukemia cells (L1210/0) and human T-lymphocyte cells (Molt4/C8 and CEM/0); 80% of these compounds were well-classified, as active or inactive, and only two pairs of isomeric compounds were false actives. The chloropurine derivatives were the most active compounds, especially compounds **6c**,**d**.

Introduction

At present the pharmaceutical industry is under increasing pressure to discover new drug leads in a faster and more efficient way than in the past. In this sense, the rational drug design strategies, especially the computer-aided drug design approaches, have emerged as a promising solution to the efficient search for new *lead compounds*.[1−3] One of the major goals of computer-aided drug design and discovery strategies is the identification of small *subsystems* from a large group of chemical compounds. A subsystem is understood to be a number of compounds formed by a significant variation in a given parent structure, which is referred to as the lead compound. These subsystems can be obtained through *virtual screening*[4] of large databases or libraries of candidate compounds. In this sense, the virtual screening is understood to be a filtering of a database through the use of a computational approach based on a discrimination function that permits the selection of a series of compounds to be assayed for biological activity.[4−6]

The search for anticancer compounds has always been on the desktop of medicinal chemists, and a great number of different approaches have been used in this search.[7−10] However, the discovery of selective antitumor compounds has remained a largely elusive goal of cancer research. Consequently, novel approaches are needed for the efficient search for novel candidates to be assayed as anticancer drugs.[11,12]

Recently, a novel scheme to the rational design and selection of molecules has been introduced by one of the present authors. It is the so-called *TO*pological *S*ub-*S*tructural *MO*lecular *DE*sign (TOSS-MODE).[13] This approach is very useful for the selection of novel subsystems of chemicals having a desired property. These subsystems can be further optimized by using some of the many molecular modeling techniques at the disposition of the medicinal chemists.

In this work the TOSS-MODE strategy to find a quantitative model that discriminates anticancer compounds from the inactive ones was used. A simulated experiment of virtual screening for the search of inhibitors of Ras FTPase is conducted by using the model developed here. Finally, the design, synthesis, and biological evaluation of a new series of carbonucleosides with antitumor activity are presented.

Results and Discussion

The TOSS-MODE Approach. The TOSS-MODE approach is based on the calculation of the spectral moments of the topological bond matrix.[13−18] First, the molecular structures of compounds under study are represented by their molecular graphs and the bond matrices[19] are then generated for these molecular graphs. This matrix, $\mathbf{B}(G)$, is a square and symmetric one, representing the adjacency between the pairs of bonds in the molecule.[19] Consequently, the nondiagonal entries of $\mathbf{B}(G)$ are ones or zeros regardless of whether the corresponding bonds in the molecule are adjacent or not. Two bonds are adjacent if they have a common atom. In two of our previous papers we were concerned only with the spectral moments of the edge adjacency matrix of hydrocarbons for which it was not necessary

* Corresponding author. Tel: 34 981 5631000, ext. 14936. Fax: 34 981 594912. E-mail: estrada66@yahoo.com.
† Universidad de Santiago de Compostela.
‡ Universidad Central de Las Villas.
§ Katholieke Universiteit Leuven.

10.1021/jm991172d CCC: $19.00
Published on Web 04/22/2000

to make any differentiation for the entries of such matrix. However, when we are concerned with molecules containing heteroatoms, an approach based on the use of diagonal weights in the bond matrices has been introduced. These weights correspond to different bond properties for organic molecules. For instance, bond lengths were used for describing antifungal activity of benzyl alcohols and boiling points and diamagnetic susceptibilities of alkyl halides. Here the standard bond dipoles taken from Ferguson[20] are used as bond weights. The use of these bond weights was first introduced for the discrimination of the sedative/hypnotic activity of a large data set of organic molecules.[13] In this previous work we showed that this magnitude produced better discrimination than the standard bond distances for the case of large data sets. The selection of the standard dipole moments compiled by Ferguson[20] to be used in the present work responds to the greater variability in these values as compared to those which were previously used by us.[13] However, a comparison of the discriminative power of the spectral moments obtained from both sets of dipole moments has not been carried out yet and will be the objective of a forthcoming work.

The kth spectral moment of the bond matrix, $\mu_k = \mu_k(G)$, is computed as the sum of the diagonal entries of the kth power of $\mathbf{B}(G)$. These spectral moments are used as molecular descriptors for finding a quantitative model of the following form:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4 + ... + a_k\mu_k + b \quad (1)$$

where P is the biological property, μ_k is the kth spectral moment, and the a_k's are the coefficients obtained by the statistical approach used, e.g., the linear discriminant analysis (LDA).

The moments of the $\mathbf{B}(G)$ matrix can be expressed as linear combinations of molecular substructures.[13−18] These contributions can be obtained via algebraic mathematical expressions or through a computational approach. The first approach generates the contributions to the biological activity of all structural fragments that contribute to the corresponding spectral moments. The second approach consists of calculating the contribution of only some selected fragments to the spectral moments. The last method is preferred for higher-order spectral moments, such as μ_i, $i \geq 10$, because of the great number of fragments contributing to them, which produces a combinatorial explosion of the fragments that need to be generated by the computer. By using this approach we can determine how a given substructure influences the biological activity under study. The second approach will be used in the present work in order to compute the contribution of different structural fragments to the anticancer activity. A methodological explanation of the calculations concerned with the TOSS-MODE approach is provided in a recent work together with several examples of its application.[21]

It is well-known that the stereochemical features of molecules are of great importance for the explanation of the biological activity of chemicals, and the fact that graph-theoretical schemes do not include explicitly these features has been considered as a limitation of such approaches.[22] However, when we are interested in searching for structural molecular patterns responsible for the biological activity in large databases of chemicals, the 2D topological approaches have proved to be very effective. In this sense 2D fingerprints and large series of topological indices are currently in use for the studies of virtual screening and molecular diversity analysis.[23] The present approach can be considered in this direction as a tool for the virtual screening of lead compounds in databases of molecular structures and as the first step in the design of lead compounds. The a posteriori consideration of the 3D features of these molecules so designed will be useful in optimizing their biological activities. For instance, Estrada[24] has recently taken into account the 3D structural features of small/medium-sized molecules and proteins inside the scheme of the TOSS-MODE approach by making use of the iterated line graph sequence of the molecular graphs.[25]

Another point that requires our attention relates to the necessity of considering physical properties explicitly in the development of the discrimination model. This point is directly related to the very nature of the approach that is in use. The present approach is based on the explicit characterization of the molecular structure of chemicals.[26] In this way, properties such as solubility in water, partition between different phases, etc., do not have necessarily to be taken into account explicitly in the model. They are also dependent on the molecular structure and consequently are implicitly considered when structure-explicit approaches, such as the present one, are used.

Developing the Classification Function. The development of a discriminant function that permits the classification of chemicals as active or inactive is the key of the present approach for the design and discovery of anticancer compounds. Thus, it is necessary to select a training data set of anticancer compounds containing a great structural variability. The quality of the classification model depends in a great percentage on the quality of the training data set selected.

Here we consider a general data set of 293 organic compounds, 86 of them having anticancer activity reported and the rest being inactive.[27] The anticancer compounds are those reported in the Negwer handbook[27] as being used in clinics for the treatment of any class of cancer. On the other hand, the inactive compounds were selected from several drugs having other uses in clinics. By this way, the active compounds of the training and prediction data set possess great structural variability and different mechanisms of anticancer activity. In this way, the selection/design process is not constrained only to lead compounds with one specific mechanism of action but also for the discovery of leads with novel mechanisms of anticancer activity. This is well-illustrated later in a simulated virtual screening experiment.

This data set was divided into two subsets: the first containing 222 compounds with 63 actives and 159 inactives used as a training set; the other set of 71 compounds with 23 actives and 48 inactives was left for the cross-validation of the model developed. The selection was carried out at random, and compounds in the prediction set were never used in the development of the classification model. The following classification model was obtained with the use of the LDA technique

Table 1. Results of the Classification of Compounds in the Training and Prediction Sets

compd	prob	class	compd	prob	class	compd	prob	class
			Training Active Group					
timonacic	76.07	+	enterodiol	12.33	−	doxifluridine	95.21	+
mercaptopurine	53.33	+	aminopterin	87.60	+	fluorafur	96.19	+
NSC-143019	20.37	−	pafencil	51.48	U	forfenimex	8.35	−
blevidon	88.23	+	trimetrexate	58.77	+	EMP	89.01	+
mitobronitol	15.26	−	testolactone	50.26	U	tiazoofurine	70.68	+
trichlormethine	55.17	+	phenaline	94.88	+	tretamine	63.70	+
trimelaminum	76.47	+	fenamet	99.88	+	chlorethylaminouracil	95.97	+
bromacrylide	24.86	−	fluoasquin	72.59	+	thyminalkylamine	88.53	+
tegafur	69.79	+	chlorasquin	73.33	+	damvar	71.12	+
uramustine	88.53	+	acodazole	35.62	−	calci mefolinas	89.29	+
6-azauridine	88.97	+	edikron	35.09	−	AT-1727	99.72	+
imidazole mustard	92.72	+	AC 3579	54.07	+	aphicolin	4.06	−
azacitidine	93.24	+	bremfol	84.21	+	KW 2083	97.76	+
SA 291	99.64	+	methopterine	80.40	+	methasquin	71.14	+
azathioprine	81.94	+	A-ninopterin	85.58	+	denopterin	80.42	+
fluoroocitabine	93.09	+	CB-10252	33.88	−	fenastezin	61.85	+
dimedazol	1.21	−	AT-16	89.56	+	promicil	47.62	U
broxuridine	91.78	+	citosal	40.89	−	A-denopterin	82.27	+
			Training Inactive Group					
dibromsalicylamide	92.76	−	clozapine	87.62	−	furazonal	85.16	−
basedol	96.47	−	prothixene	99.89	−	brodimoprim	58.66	−
metipirox	95.82	−	methonalide	87.17	−	selectan	99.27	−
chlorphenesin	91.10	−	prothiendyl	99.51	−	brosotamide	95.69	−
bromaspirin	80.73	−	dehydroclothepine	98.97	−	milacemide	99.04	−
metofurone	52.20	U	pacilan	34.61	+	chlorphenacemide	89.30	−
nitrofurylen	92.39	−	oxanamide	98.37	−	atrolactamide	99.10	−
BT132 merek	98.46	−	pentabamate	61.98	−	cinromide	99.33	−
diampromide	96.28	−	PF-257	98.66	−	zebromal	82.69	−
metacetanilidum	99.61	−	metaclazepam	91.88	−	phensuximide	86.22	−
etenzamide	99.30	−	fenharmane	95.60	−	TG 38	93.46	−
parapropamol	99.11	−	diperazinepentanol	81.32	−	benzamidesalicylic acid	95.44	−
thiocylate	98.89	−	oxazepam	88.30	−	orthovanizide	83.10	−
alminoprofen	90.89	−	difenclozazine	88.42	−	vadrine	90.88	−
AP-752	52.53	U	mebicar	4.32	+	strinoline	99.38	−
methamilane	94.16	−	salinazid	94.09	−	lofemizole hydrochloride	99.20	−
metacetamol	98.76	−	bromocyl	98.58	−	B.W. 775C	98.00	−
acetaminophen	98.85	−	PAMSA	97.55	−	alclofenac	82.24	−
furaguanidine	96.09	−	T_{28}	95.87	−	emorfazone	16.19	+
diamide	80.37	−	diphenizin	98.20	−	diclonixin	80.40	−

as implemented in the STATISTICA computer software:[28]

$$C = 1.234\mu_1 - 0.552\mu_2 - 0.242\mu_3 + 0.622\mu_4 - 0.125\mu_6 + 0.022\mu_7 + 7.27 \times 10^{-4}\mu_8 - 4.11 \times 10^{-7}\mu_{14} + 6.782 \times 10^{-8}\mu_{15} - 0.040\mu_0\mu_1 + 6.949 \times 10^{-6}\mu_0\mu_7 - 9.998 \quad (2)$$

$$N = 224,\ \lambda = 0.557,\ D^2 = 3.90,\ F(11,212) = 15.33$$

The Wilks' λ statistic for the overall discrimination can takes values in the range of 0 (perfect discrimination) to 1 (no discrimination). The Mahalanobis distance indicates the separation of the respective groups. It shows whether the model possesses an appropriate discriminatory power for differentiating between the two respective groups. In developing this classification function the values of 1 and −1 were assigned to active and inactive compounds. By using this model one compounds is classified as active if $C \geq 0.2$ and inactive if $C \leq -0.2$. The interval between −0.2 to 0.2 is left for not classified compounds, see below. However, we will use the a posteriori probabilities instead of these cutoff values in order to classify the compounds as active/inactive. This is the probability that the respective case belongs to a particular group (active or inactive). It is proportional to the Mahalanobis distance from that group centroid. This model classified correctly 88.29% of compounds in the training series, i.e., 26 misclassifications in 222 cases, while in the prediction set there were 6 errors in 71 cases, that is, 91.55% of good classification. The percentage of false actives in the training set was only 4.95% (11 inactive compounds were classified as actives from 222 cases), and 6.76% were false inactives (15 anticancer compounds were classified as inactive). These percentages for the prediction set are only 1.43% of false actives and 7.14% of false inactives. It is desirable that the number of false active compounds be as low as possible because this number represents inactive compounds that will be sent to the biological assays with the consequent loss of time and resources.

In Tables 1 and 2 the results of classification and a posteriori probabilities for some compounds, in both training and external prediction sets, are shown. The complete set of compounds in training and prediction sets as well as their classification is given as Supporting Information. In the training series there are 10 compounds for which the percentage of classification as active and inactive does not differ by more than 5%. These compounds are considered as unclassified and are designated by U in Tables 1 and 2. In the cross-validation set there was only one unclassified compound. A graphical visualization of the results obtained for the

Table 2. Results of the TOSS-MODE Classification of Compounds in an External Prediction Set

compd	prob	class	compd	prob	class
Test Active Group					
ancitabine	86.33	+	chlormethine	13.25	−
caracemide	48.88	U	fluorodopan	95.57	+
aethimidinum	26.23	−	cytarabine	87.06	+
cervicarcin	68.05	+	butotricin	95.81	+
1954 C.B.	92.14	+	hisfen	94.50	+
crotepoxide	82.61	+	ambunol	36.62	−
pretazetine	95.60	+	quinaspar	62.66	+
quinocarcin	95.60	+	methotrexate	82.09	+
leucenol	40.70	−	ZIMET 54/79	63.32	+
camptothecin	96.32	+	blenfuron	1.21	−
vinervine	98.33	+	lofenal	61.14	+
leucenol	40.70	−	dimetfolamide	80.42	+
floxuridine	91.94	+	TPN351	99.91	+
Test Inactive Group					
mycosid	93.90	−	dihalopromazine	96.82	−
aesulanum	99.59	−	methopromazine	99.40	−
allisan	96.59	−	fluperlapine	61.03	−
triladine	78.09	−	valnoctamine	98.09	−
salicilamid	99.69	−	loprodiol	94.61	−
paracetamol	73.32	−	mecloralurea	92.57	−
acidum nicosalicylicum	91.22	−	PAS-hydrazide	99.09	−
acylnidrazone	83.33	−	metazide	63.41	−
SOG-18	90.90	−	isoniazid	98.99	−
brosuximide	74.71	−	brofezil sodium	95.83	−
SL 75102	96.98	−	praxadine	99.21	−
ethosuximide	90.96	−	pirprofen	75.84	−
contidine fumarate	85.81	−	GOBAB	97.24	−
etoloxamide	98.50	−	diclofenac sodium	96.66	−
azamianserin	84.73	−	KB 1043	78.05	−

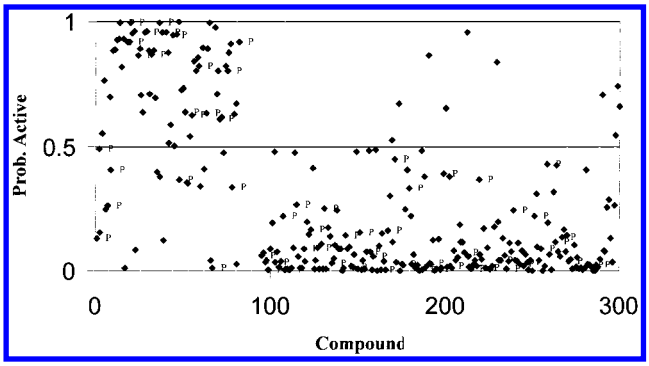


Figure 1. Plot of the probability of active for each compound in both the training and prediction sets (P). Active compounds are those with numbers below 100.

discrimination of active/inactive compounds in both the training and prediction series is given in Figure 1.

Search for Active/Inactive Fragments. One of the principal advantages of the TOSS-MODE approach is the possibility of obtaining the contributions of different structural fragments to the biological activity studied.[13] By using the approach previously explained we compute the contribution to the anticancer activity of a series of molecular substructures from a database of fragments previously elaborated. The quantitative contributions of fragments to the anticancer activity should be understood as the contribution that each substructure makes to the discrimination function (model 2). Hence, the qualities of these numerical values are the same as the qualities of the discrimination function. The most important aspect of these contributions is that they can serve, for medicinal chemists, as a guide to the synthesis of novel compounds including these substructures, or to the detection of possible 2D pharmacophores. The criterion which we are applying here is that of the utility of such numerical results, in complete agreement with the statement of Craig[29] that "substructures are man-made concepts, of varying size, whose definition is arbitrary, and the utility of a given substructure is the sole measure of its validity".

These fragments included some substructures present in known anticancer compounds, as well as in inactive compounds. In Figure 2 we illustrate the quantitative contribution of some of the selected fragments. This list of structural fragments is by no means exhaustive but represents an illustration of the possibilities that the present theoretical approach brings to medicinal chemists. It is clear that the presence of fragments with positive contributions to the activity in a molecule is not a requisite for the development of anticancer action. It is well-known that the biological activity is the result of the effects produced by all the structural fragments and the molecule as a whole. These active fragments, however, can lead medicinal chemists to important insights about the groups that they need to introduce or avoid (in the case of negative contributions) in order to increase the probability of finding active compounds. On the other hand, important insights about the 2D pharmacophores and the possibilities of bioisosteric substitution of groups are also obtained from the analysis of the contributions of fragment to the biological activity.[13]

It is interesting to note that there is some resemblance in the contributions to the anticancer activity obtained with the present approach and other results reported through the use of very different approaches. For instance, some coincidences exist between some of the active fragments reported by Chu et al.,[30] and those reported here. These authors used the number of fragments of different types as variables for discriminating anticancer compounds from inactive ones in a substructural approach to drug design. They found, for instance, positive contributions for pyridine ring, oxygen-containing cycles, urea group, and the groups formed by two nitrogen (or oxygen) atoms separated by two carbons. We give similar results in Figure 2.

Virtual Screening of Ras FTase Inhibitors. The virtual screening has emerged as an interesting alternative to high-throughput screening.[4−6] It mainly consists of the handling and screening of large databases in order to find a reduced set of compounds for which a prediction of a specific biological activity has been made using clustering and similarity searching. After this, other molecular modeling techniques, such as docking and molecular superposition, are used to discover pharmacophores and to refine the designs. Here we simulate a virtual search of Ras farnesyltransferase (FTase) inhibitors by using the discriminant function obtained through the TOSS-MODE approach.

To avoid the manipulation of large databases of chemicals we have selected a series of compounds, previously reported as Ras FTase inhibitors, that will be evaluated by model 2 as active/inactive ones. This data set includes compounds of very different chemical families of natural or synthetic origin. However, the most challenging aspect of this simulation experiment is not the great structural variability of the data set of compounds but the fact that for developing model 2 no one Ras FTase inhibitor was used. In this way, the evaluation of these compounds, which are completely

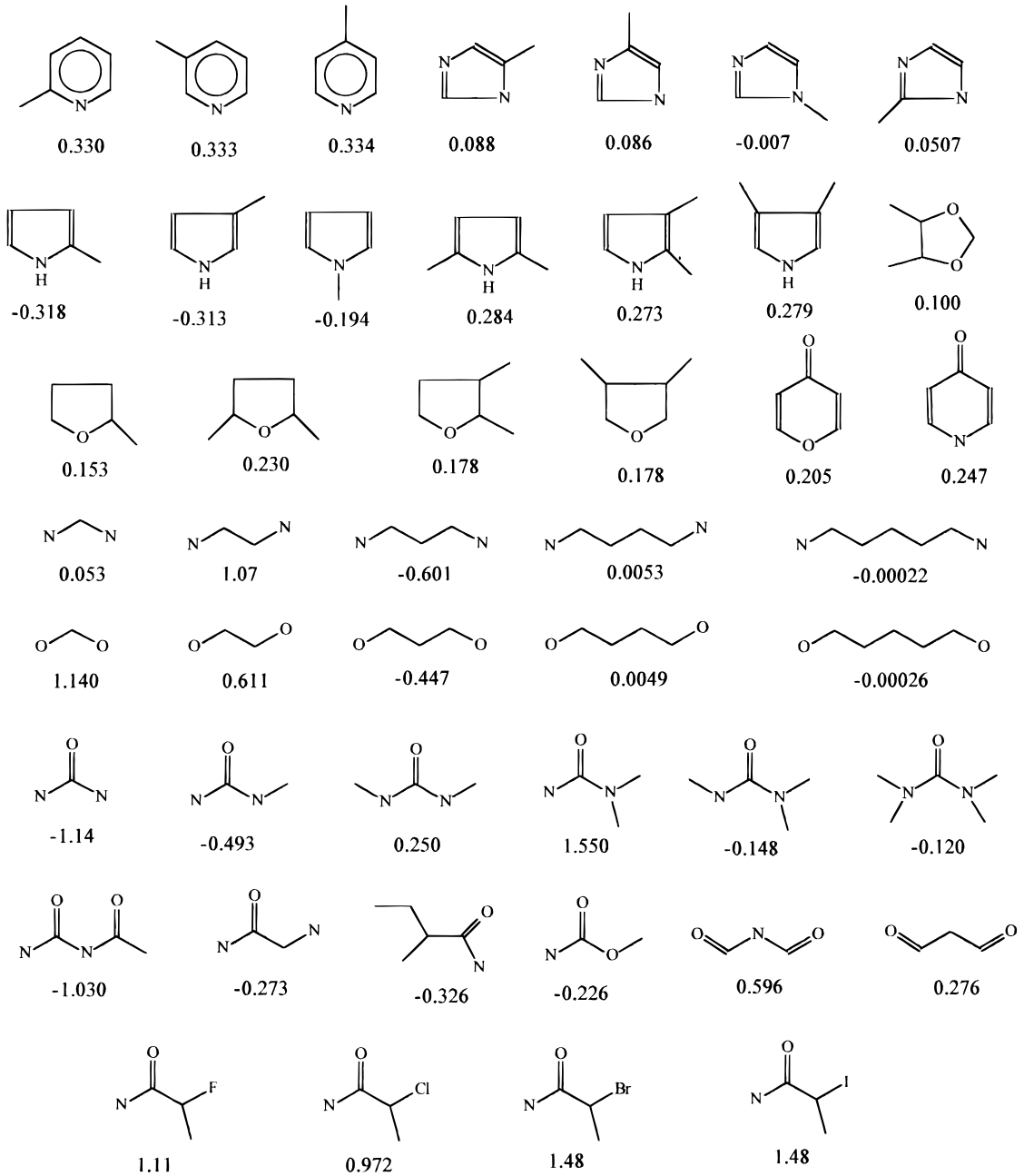


Figure 2. Molecular fragments (substructures) and their contributions to the anticancer activity according to the TOSS-MODE model obtained here.

"unknown" (from both structural and mode of action) by model 2, is equivalent to the discovery of new lead compounds with a novel action mechanism.

The FTase inhibitors may offer novel therapeutic opportunities as anticancer drugs, based on the idea that the inhibition of this enzyme can inhibit the function of Ras proteins blocking the uncontrolled mitogenic signaling pathway.[31] Therefore, the search for potent inhibitors of Ras FTase has recently become an intensive area of research in medicinal chemistry.[31,32]

In this experiment 30 compounds previously reported as Ras FTase inhibitors were evaluated through model 2. This data set of structural diverse compounds includes Ras FTase inhibitors based on the CAAX motif, constrained CAAX mimetics with mimetic replacement of AA in the CAAX and natural products.[31−34] The results of evaluating these compounds by using model 2 are given in Table 3, and the molecular structures of these compounds are illustrated in Figure 3. As can be seen in this table, most of the 30 compounds evaluated are predicted as actives. Only 4 of them are predicted to be inactive, which represents only 13% of false inactives.

It is obvious from these results that the quality of the predictions carried out for these compounds is by no means fortuitous. This is a very valuable result because, as previously stated, there was neither any FTase inhibitor nor any compound with a similar molecular structure included in the training set used in this study. An appropriate explanation might be that the TOSS-MODE approach is able to detect and discriminate the molecular structural features responsible for a specific biological activity considering implicitly the substructural and global molecular features at the same time.

Table 3. Results of the Virtual Screening Simulation of Inhibitors of Ras FTase

compd[a]	name	prob	class
a	L-731,734	99.65	+
b	L-731,735	99.25	+
c	B 581	99.96	+
d	B 956	99.97	+
	B 1086	99.96	+
e	Chart 3a in ref 31	83.13	+
f	Chart 3b in ref 31	99.52	+
g	Chart 3c in ref 31	99.60	+
h	BZA-2B	56.81	+
	BZA-5B	65.81	+
i	Cys-3-AMBA-Met	99.74	+
j	(*R*)-4-[*N*-(3-mercapto-2-aminopropyl)-amino]-3′-carboxybiphenyl	3.34	−
k	FTI-276	90.57	+
	FTI-277	90.85	+
l	L-745,631	39.05	−
m	chaematomellic acid A	22.76	−
n	chaematomellic acid B	61.31	+
o	UCF1-A	99.31	+
	UCF1-C	99.08	+
p	fusidienol	64.90	+
q	preussomerins	97.38	+
r	10′-desmethoxystreptonigrin	26.87	−
s	cylindrol A	70.31	+
t	kampanol A	94.03	+
u	kampanol B	91.50	+
v	kampanol C	94.67	+
w		99.62	+
x		99.4	+
y	SCH 56580	82.90	+
z	H-D-Trp-D-Met-D-Fcl-Gla-NH_2	86.51	+

[a] Compounds **t**−**v** were taken from ref 33, compunds **w** and **x** were taken from ref 34, and the rest are taken from refs 31 and 32.

This confirms our previous results where TOSS-MODE was able to detect novel lead compounds with sedative/hypnotic activity from a training set containing no analogous of the compounds discovered.[13]

Discovery of Novel Carbonucleoside Anticancer Compounds. In the present work a design approach for the rational selection (discovery) of novel compounds with anticancer activity has been employed. The design is conceived here as "a preliminary sketch or outline showing the main features of something to be executed". In this way, we select the family of carbonucleoside compounds to be studied with the present approach.

The design process started by drawing the chemical structures of some of the carbonucleosides obtained in one of our groups and several others not even synthesized. This type of compounds can exist as two different geometrical isomers: 1,2-cis and 1,2-trans. It is known that the present approach makes no differentiation between stereochemical features of molecules, i.e., it is a topological (2D) approach. Consequently, we consider the bi-dimensional structure of each carbonucleoside as the entry for its evaluation by the TOSS-MODE. However, for each 2D chemical structure evaluated by the TOSS-MODE we need to assay two isomeric compounds for their anticancer activity. Thus, from the point of view of evaluating the quality of TOSS-MODE predictions, a compound predicted as active is not considered a false active if at least one of both geometrical isomers shows the anticancer activity in vitro.

The compounds so designed were evaluated by the TOSS-MODE model and then synthesized by following the strategy given in Scheme 1. As shown in Scheme 1, compounds **3**, **4**, and **6**−**8** were synthesized starting from the corresponding amino alcohols of type **1**, which were obtained by different alternative forms, from commercially available ethyl 2-oxocyclopentylcarboxylate.[35]

The compounds **3** and **4** were obtained from the corresponding amino alcohols **1**, by treatment with 3-ethoxy-2-propenoyl isocyanate[36] in dimethylformamide at room temperature,[37] followed by cyclization of the resulting disubstituted urea (compounds **2**) by refluxing sulfuric acid medium (65−70% overall yield). Compounds of type of **4** were synthesized in 90−95% yield by treatment of the corresponding compounds of type **3** with iodine−nitric−acid in dioxane.[38]

To obtain the purine derivatives **6**−**8**, the amino alcohols **1** were condensed with 5-amino-4,6-dichloropyrimidine in refluxing *n*-butanol containing triethylamine,[39] providing the corresponding compounds **5** in 85−90% yields, and ring closure with ethyl orthoformate in acid medium giving the 6-chloropurines **6** in 80% quantitative yields. The compounds of type **6** were treated with a solution of ammonia in methanol to provide the adenine derivatives **7** or with hot sodium hydroxide[40] to furnish the hypoxanthine derivatives **8** in nearly quantitative yields.

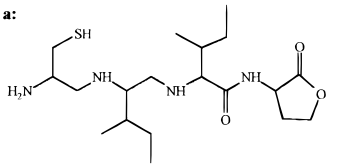
The compounds **3**, **4**, and **6**−**8** were previously classified as active/inactive according to their anticancer activity by using model 2 obtained in a preceding section. After this, 20 compounds classified as active or inactive were evaluated experimentally for their antitumor effects on the proliferation of murine leukemia cells (L1210/0) and human T-lymphocyte cells (Molt4/C8 and CEM/0).[36] The results of both classification and experimental pharmacological evaluations are given in Table 4.

As can be seen in Table 4 three pairs of isomeric compounds were theoretically classified as inactive ones. They were compounds **3a**,**b**; **3c**,**d**; and **7a**,**b**. The compounds of the type **3** became inactive in inhibiting the proliferation of the three cellular lines used, and they can be considered as inactive ones, in complete agreement with the theoretical predictions. The compound **7b** showed some activity against two cellular lines; however, the values of IC_{50} for this compound are very high (the highest of all active compounds) demonstrating that this compound has a poor antitumor activity.

Most interesting are the results of active compounds, consisting of seven pairs of isomeric compounds that were predicted to be active for model 2. Three of these pairs of compounds were active for both isomers, i.e., cis and trans. In another two pairs of compounds at least one of the two isomers became active in the experimental evaluation and thus were considered as well-classified. The other two pairs of compounds resulted inactive in the experiment and are considered as false actives.

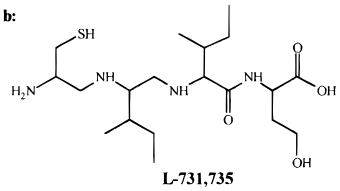
The two pairs of false active compounds represent 20% of bad classification of the complete data set of carbonucleosides evaluated. These compounds, **8a**,**b** and **8c**,**d**, are very similar (from a topological and geometrical point of view) to compounds of type **7**. The only difference between them is the change of the NH_2 group in **7** to an OH in **8**. However, it appears from the experimental values of the antitumor activity that this change is enough for the suppression of the activity in

a:



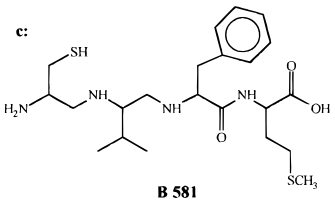
L-731,734

b:



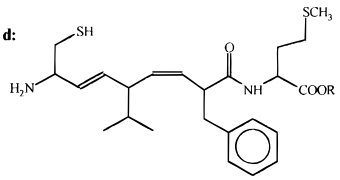
L-731,735

c:



B 581

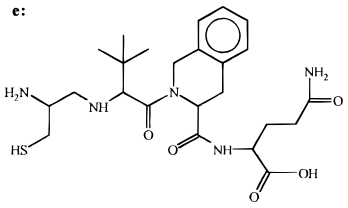
d:



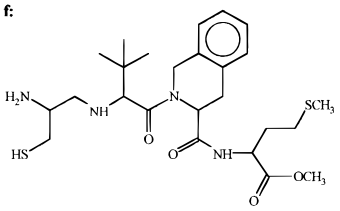
B 956: R = H

B 1086: R = CH_3

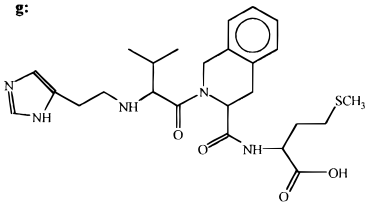
e:



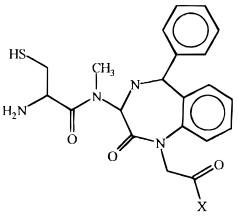
f:



g:



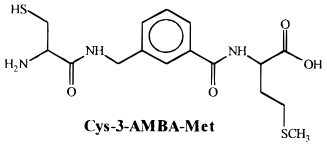
h:



BZA-2B: X = CH_2-OH

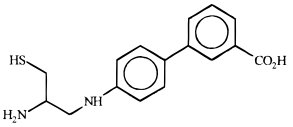
BZA-5B: X = CH_2-OCH_3

i:

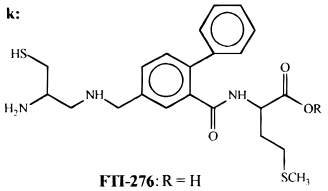


Cys-3-AMBA-Met

j:



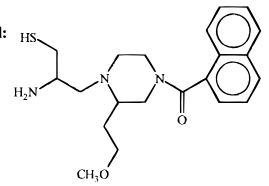
k:



FTI-276: R = H

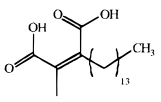
FTI-277: R = CH_3

l:



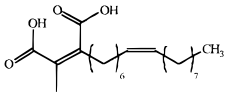
L-745,631

m:



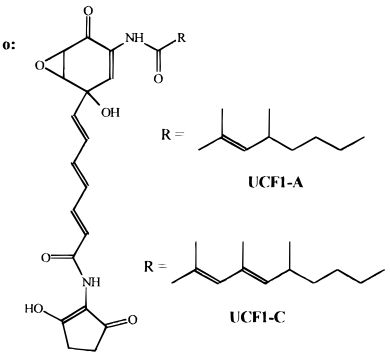
Chaetomellic Acid A

n:



Chaetomellic Acid B

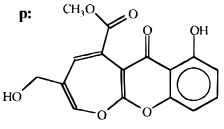
o:



UCF1-A

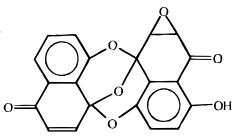
UCF1-C

p:



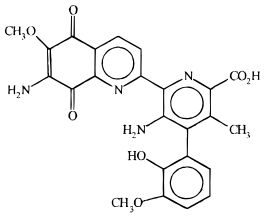
Fusidienol

q:

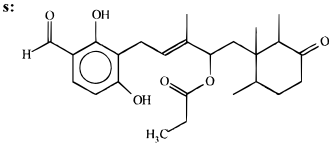


Preussomerins

r:

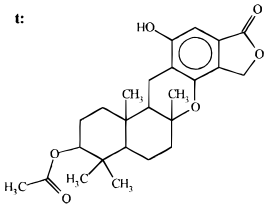


s:



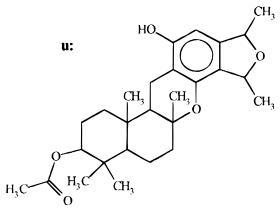
Cylindrol A

t:



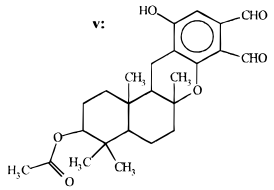
Kampanol A

u:



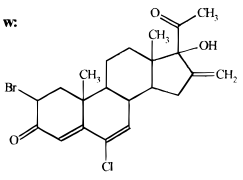
Kampanol B

v:

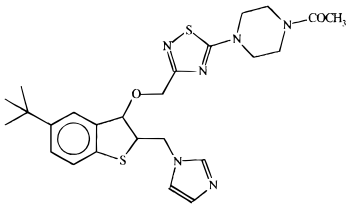


Kampanol C

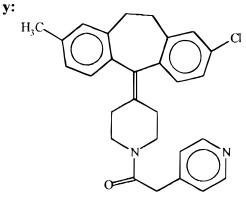
w:



x:

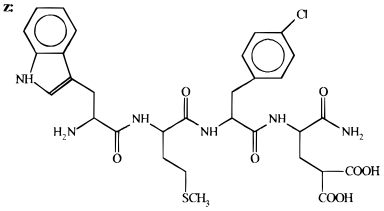


y:



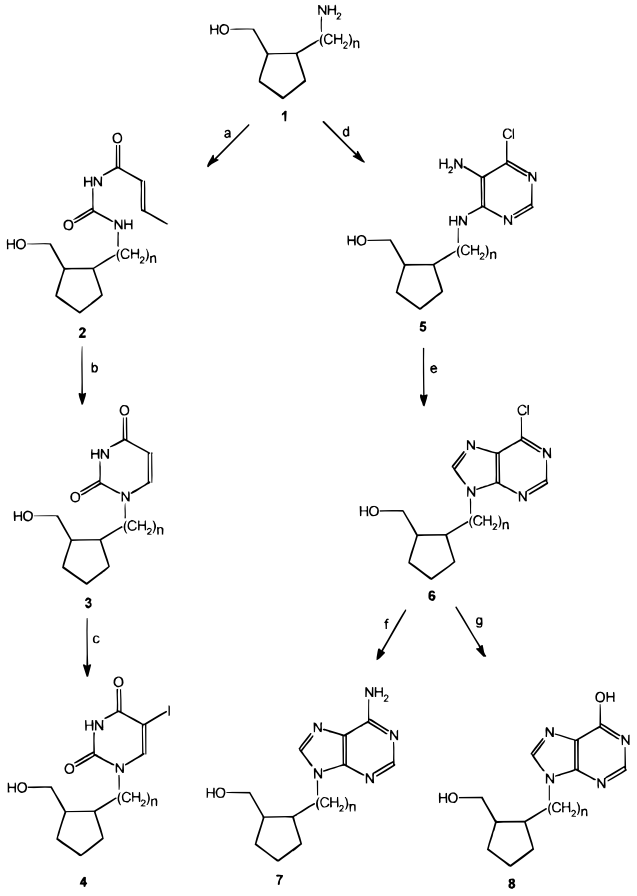
SCH 56580

z:



H-D-Trp-D-Met-D-Fcl-Gla-NH_2

Scheme 1[a]



1a n = 0 cis **2a** n = 0 cis **3a** n = 0 cis **4a** n = 0 cis
1b n = 0 trans **2b** n = 0 trans **3b** n = 0 trans **4b** n = 0 trans
1c n = 1 cis **2c** n = 1 cis **3c** n = 1 cis **4c** n = 1 cis
1d n = 1 trans **2d** n = 1 trans **3d** n = 1 trans **4d** n = 1 trans

5a n = 0 cis **6a** n = 0 cis **7a** n = 0 cis **8a** n = 0 cis
5b n = 0 trans **6b** n = 0 trans **7b** n = 0 trans **8b** n = 0 trans
5c n = 1 cis **6c** n = 1 cis **7c** n = 1 cis **8c** n = 1 cis
5d n = 1 trans **6d** n = 1 trans **7d** n = 1 trans **8d** n = 1 trans

[a] Reagents: (a) 3-ethoxy-2-propenoyl isocyanate, DMF, benzene; (b) H_2SO_4; (c) I_2, HNO_3, dioxane; (d) 5-amino-4,6-dichloropyrimidine, Et_3N, *n*-BuOH; (e) $CH(OEt)_3$, HCl; (f) NH_3, MeOH; (g) NaOH.

the case of compound **8c** compared to its analogue **7c**. This difference is not well-accounted for by the present approach, and the introduction of some molecular descriptor is necessary for a better discrimination between these two sets of compounds.

According to the criterion of good classification (based on the nonconsideration of stereochemical differences) 80% of carbonucleosides were well-classified. From the compounds predicted as active the chloropurines were the most potent inhibitors of the proliferation of the three cellular lines, i.e., compounds of type **6**, especially compounds **6c**,**d**. It is not clear at present the nature of the structure−activity relationships in this series of carbonucleosides. In general cis compounds are more active than trans isomers, probably due to their greater 3D similarity with the natural nucleosides. However, there are some cases in which the inverse order is observed, e.g., compounds **6a**,**b**. Similarly, the influence of the length of the chain separating the cyclopentane ring and the nucleoside base does not follow a general rule. In chloropurine compounds the presence of a CH_2 group separating the cyclopentane ring and the purine favors the antitumor activity for both cis and trans isomers. In other words, these compounds are more active than their analogues in which both rings are directly bonded. However, in the adenine (**7**) and uracil (**4**) compounds no regularity exists. It is clear that more data is necessary in order to obtain clearer structure−activity relationships for these series of compounds. The main objective of the present work is to only show the possibilities of the TOSS-MODE approach in the search for novel anticancer compounds. Further studies by using this approach in the formulation of structure−anticancer activity of this type of carbonucleoside are now in progress by our research groups. However, the findings reported here demonstrate that the TOSS-MODE approach is useful in discriminating compounds with anticancer activity from those being inactive.

Table 4. Carbonucleoside Compounds Designed in the Present Study, Their Classification (a posteriori probabilities) According to the TOSS-MODE Approach, and Their Anticancer Activity against Three Cellular Lines

			IC_{50} (mg/mL)[a]		
compd	prob	class	L1210/0	Molt4/C8	CEM/0
3a	22.9	−	>200	>200	>200
b			>200	>200	>200
c	23.1	−	>200	>200	>200
d			>200	>200	>200
4a	65.1	+	52.3 ± 1.9	77.4 ± 4.8	57.4 ± 0.4
b			>200	>200	>200
c	65.6	+	71.1 ± 7.5	108 ± 38	117 ± 7
d			65.8 ± 4.4	158 ± 59	131 ± 48
6a	59.5	+	62.8 ± 04	49.9 ± 2.5	61 ± 7
b			56.1 ± 3.0	39.2 ± 8.0	19 ± 4
c	62.4	+	14.0 ± 1.1	6.54 ± 0.47	16 ± 1
d			48.7 ± 0.8	22.3 ± 16.7	15 ± 2
7a	34	−	>200	>200	>200
b			179 ± 20	>200	180 ± 28
c	57.5	+	102 ± 3	100 ± 18	108 ± 11
d			>200	>200	>200
8a	62.9	+	>200	>200	>200
b			>200	>200	>200
c	60.3	+	>200	>200	>200
d			>200	>200	>200

[a] 50% Inhibitory concentration, or compound concentration required to reduce proliferation of tumor cells by 50%.

Concluding Remarks

We have shown here that the TOSS-MODE approach is able to produce adequate models to correctly classify the anticancer activity of several types of organic compounds. This approach, based on simple graph-theoretical concepts, permits the identification of active/inactive fragments that can be used as guides for the design of novel compounds and for the search of possible pharmacophores. The simulated experiment of virtual search of Ras FTase inhibitors has proved the capacity of this approach in the identification in databases of lead compounds having novel chemical structures and new modes of action.

The discovery of novel compounds with anticancer activity has proved that the TOSS-MODE permitted the correct classification of 80% of a new series of carbo-

Figure 3. Molecular formulas of the Ras FTase inhibitors evaluated by the TOSS-MODE model in a virtual screening experiment.

nucleosides. One of these compounds, regarded as active by the model developed in this work, has shown an interesting profile of activity against the proliferation of murine leukemia cells (L1210/0) and human T-lymphocyte cells (Molt4/C8 and CEM/0).

The TOSS-MODE approach for the search of anticancer compounds can be considered as an iterative strategy in the sense that the compounds that are discovered in the search of novel anticancer drugs can be incorporated into the training series and a new classification model can then be obtained. This approach will guarantee that the molecular structures of novel lead compounds are represented in the training series, thus permitting a better discrimination of this type of compounds and the identification of novel structural patterns for the desired activity. For example, the structures of the novel carbonucleosides obtained in this research can be included into a new training data set for developing a classification model with improved capacity for discriminating this type of compound. Some work in this direction is now in progress and will be published in a forthcoming paper.

Experimental Section

Chemistry. Melting points are uncorrected and were determined in a Reichert Kofler thermopan or in capillary tubes in a Buchi 510 apparatus. IR spectra were recorded in a Perkin-Elmer 1640FT spectrometer (KBr disks, υ in cm^{-1}). ^1H NMR (300 MHz) and ^{13}C NMR (75.4 MHz) spectra were recorded in a Bruker AMX spectrometer, using TMS as internal standard (chemical shifts in δ values, J in Hz). Mass spectrometry was carried out on a Kratos MS-50 or on a Varian MAT-711 spectrometer. Elemental analyses were performed by a Perkin-Elmer 240B microanalyzer and were within ±0.4% of calculated values in all cases. Flash chromatography (FC) was performed on silica gel (Merck 60, 230−400 mesh); analytical TLC was performed on precoated silica gel plates (Merck 60 F254, 0.25 mm).

(±)-*trans*-1-[2-(Hydroxymethyl)cyclopentyl]uracil (3b). A solution of a 50% mixture of (±)-2-(hydroxymethyl)cyclopenthylamine *cis* (**1a**) and *trans* (**1b**) (1.34 g, 10 mmol) in dimethylformamide (35 mL) was stirred at −20 °C over 4 Å molecular sieves and under an Ar atmosphere. A benzene solution of 3-ethoxy-2-propenoyl isocyanate[41] (29.6 mL, 10 mmol) was added, and the mixture was allowed to reach room temperature while stirring overnight. The molecular sieves were filtered and the solvent was evaporated under reduced pressure, maintaining the temperature below 40 °C (toluene and ethanol were added so as to form a low-boiling ternary azeotrope). The solid residue was purified by FC with chloroform as eluent, to obtain the mixture of (±)-3-ethoxy-*N*-[2-(hydroxymethyl)cyclopentylcarbamoyl]-2-propenamide *cis* (**2a**) and *trans* (**2b**) (2.33 g, 78%).

A solution of the mixture of the cis and trans isomers **2a**,**b** (2.33 g, 9.10 mmol) in 2M H_2SO_4 (70 mL) was refluxed 3 h. The mixture was allowed to cool, and then the solvent was evaporated under reduced pressure. The residue was purified by FC with 98:2 CH_2Cl_2/CH_3OH as eluent, which gave pure **3a**[42] (857 mg, 46%) and **3b** (860 mg, 46%): mp 190−191 °C. ^1H NMR (DMSO-d_6): 10.80 (bs, 1H, NH), 7.68 (d, 1H, H-6, J = 8.00), 5.55 (d, 1H, H-5, J = 8.00), 4.53 (t, 1H, OH, J = 4.95), 4.45 (m, 1H, CH−N, J = 8.40), 3.32 (m, 2H, CH_2−O), 2.14 (m, 1H, CH−C−O), 1.93−1.34 (m, 6H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 163.4, 151.3, 143.3, 101.4, 62.8, 59.0, 45.8, 30.9, 27.5, 22.5. IR: 3345, 3035, 1718, 1686, 1485, 1388, 1279, 797. MS *m*/*z* (%): 210 (M^+, 17), 192 ($M^+ - CH_3O$, 14), 113 ($M^+ - C_6H_9O$, 100), 112 ($M^+ - C_6H_{10}O$, 18), 81 (25), 80 (32), 67 (28). Anal. ($C_{10}H_{14}N_2O_3$) C, H, N.

(±)-*trans*-1-[2-(Hydroxymethyl)cyclopentyl]-5-iodouracil (4b). To a solution of uracil derivative **3b** (60 mg, 0.28 mmol) and 0.75 M nitric acid (0.40 mL) in dioxane (3 mL), I_2 (140 mg, 0.57 mmol) was added, and the mixture refluxed 30 min. Then it was neutralized with 0.5 M NaOH, and the solvent was evaporated under reduced pressure. The residue was purified by FC with 99:1 CH_2Cl_2/CH_3OH as eluent, which gave pure **4b** (90 mg, 95%): mp 224−225 °C. ^1H NMR (DMSO-d_6): 11.56 (s, 1H, NH), 8.13 (s, 1H, H-6), 4.52 (t, 1H, OH, J = 4.90), 4.41 (m, 1H, CH−N, J = 8.50), 3.30 (m, 2H, CH_2−O), 2.23 (m, 1H, CH−C−O), 1.89−1.22 (m, 6H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 160.6, 150.8, 147.6, 68.7, 62.8, 60.1, 45.2, 30.6, 27.1, 22.1. IR: 3028, 1689, 1655, 1602, 1452, 1271. MS *m*/*z* (%): 336 (M^+, 24), 239 ($M^+ - C_6H_9O$, 81), 238 ($M^+ - C_6H_{10}O$, 88), 195 (52), 97 (63), 81 (100), 69 (98), 67 (62). Anal. ($C_{10}H_{13}IN_2O_3$) C, H, N.

(±)-*trans*-1-[2-(Hydroxymethyl)cyclopentylmethyl]-5-iodouracil (4d). This compound was prepared from **3d** (100 mg, 0.45 mmol) in an analogous way to **4b** but refluxing the reaction mixture for 3 h. The crude product was purified by FC with 3:7 hexane/ethyl acetate as eluent, which gave pure **4d** (140 mg, 89%): mp 146−148 °C. ^1H NMR ($CDCl_3$): 10.02 (bs, 1H, NH), 7.71 (s, 1H, H-6), 3.98−3.47 (m, 5H, CH_2−O, CH_2−N + OH), 2.17−1.25 (m, 8H, CH−C−N + CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 161.0, 151.5, 149.8, 67.9, 66.6, 53.9, 45.3, 43.4, 31.5, 29.2, 24.8. IR: 3447, 1716, 1646, 1439, 1353, 1023, 610. MS *m*/*z* (%): 350 (M^+, 3), 239 ($M^+ - C_7H_{11}O$, 61), 238 ($M^+ - C_7H_{12}O$, 53), 95 (59), 94 (100), 79 (27), 67 (20). Anal. ($C_{11}H_{15}IN_2O_3$) C, H, N.

(±)-*cis*-5-Amino-6-chloro-4-[2-(hydroxymethyl)cyclopentylamino]pyrimidine (5a). A mixture of amino alcohol **1a** (100 mg, 0.87 mmol), 5-amino-4,6-dichloropyrimidine (150 mg, 0.91 mmol), triethylamine (0.5 mL) and *n*-BuOH (3 mL), was refluxed 24 h under Ar atmosphere. The solvent was then evaporated under vacuum and the solid residue dissolved in ethyl acetate by stirring with IRA-420 (OH) until all turbidity had disappeared. The resin was filtered out, the solvent evaporated under vacuum and the residue purified by FC with 98:2 CH_2Cl_2/CH_3OH as eluent, producing pure **5b** (150 mg, 71%): ^1H NMR (DMSO-d_6): 7.70 (s, 1H, H-2), 6.36 (d, 1H, NH, J = 7.40), 5.10 (bs, 2H, NH_2), 4.44 (m, 1H, CH−N), 3.25 (m, 2H, CH_2−O), 2.18 (m, 1H, CH−C−O), 1.94−1.48 (2m, 1+5H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 152.1, 145.8, 137.0, 123.8, 61.1, 53.8, 44.6, 31.7, 27.0, 21.8. IR: 3356, 3259, 2878, 1833, 1648, 1585, 1498, 1475, 1459, 1411, 1018, 1007. MS *m*/*z* (%): 244 ($[M + 2]^+$, 15), 242 (M^+, 46), 221 (39), 169 (18), 146 ($[M + 2]^+ - C_6H_{10}O$, 34), 144 ($M^+ - C_6H_{10}O$, 100), 117 (10), 67 (9). Anal. ($C_{10}H_{15}ClN_4O$) C, H, N.

(±)-*cis*-5-Amino-6-chloro-4-[2-(hydroxymethyl)cyclopentylmethylamino]pyrimidine (5c). This compound was prepared from **1c** (300 mg, 2.32 mmol) in an analogous form to **5a**. The crude product was purified by FC with 99:1 CH_2Cl_2/CH_3OH as eluent, which gave pure **5c** (525 mg, 88%): mp 148−150 °C. ^1H NMR (DMSO-d_6): 7.70 (s, 1H, H-2), 6.61 (m, 1H, NH), 5.00 (bs, 2H, NH_2), 4.43 (t, 1H, OH aliphatic, J = 4.65), 3.46 (m, 2H, CH_2−N), 3.26 (m, 2H, CH_2−O), 2.22−2.07 (2m, 1+1H, CH−C−N + CH−C−O), 1.68−1.31 (2m, 3+3H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 152.1, 145.7, 136.5, 123.3, 61.0, 43.3, 41.4, 40.6, 29.3, 27.7, 22.5. IR: 3356, 3259, 2944, 1659, 1587, 1472, 1424, 1341, 1050. MS *m*/*z* (%): 258 ($[M + 2]^+$, 5), 256 (M^+, 15), 242 ($[M + 2]^+ - NH_2$, 5), 240 ($M^+ - NH_2$, 16), 159 ($[M + 2]^+ - C_6H_{11}O$, 34), 157 ($M^+ - C_6H_{11}O$, 100), 146 ($[M + 2]^+ - C_7H_{12}O$, 33), 144 ($M^+ - C_7H_{12}O$, 98), 94 (18), 67 (10). Anal. ($C_{11}H_{17}ClN_4O$) C, H, N.

(±)-*trans*-5-Amino-6-chloro-4-[2-(hydroxymethyl)cyclopentylmethylamino]pyrimidine (5d). This compound was prepared from **1d** (500 mg, 3.87 mmol) in an analogous way to **5a**. The crude product was purified by FC with 99:1 CH_2Cl_2/CH_3OH as eluent, which gave pure **5d** (845 mg, 85%): mp 122 °C. ^1H NMR (DMSO-d_6): 7.69 (s, 1H, H-2), 6.78 (t, 1H, NH, J = 5.00), 5.00 (bs, 2H, NH_2), 4.57 (t, 1H, OH, J = 5.00), 3.42−3.23 (m, 2H, CH_2−N + CH_2−O), 1.91−1.25 (m, 8H, CH−C−N + CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 152.5, 146.0, 136.9, 123.6, 64.9, 45.9, 45.8, 41.8, 30.9, 29.4, 24.4. IR: 3233, 2942, 1607, 1593, 1462, 1420, 1346, 1219, 847. MS *m*/*z* (%): 258 ($[M + 2]^+$, 9), 256 (M^+, 26), 242 ($[M + 2]^+ - NH_2$, 6), 240 ($M^+ - NH_2$, 20), 160 ($[M + 2]^+ - C_6H_{10}O$, 12), 159 ([M

+ 2]$^+$ − $C_6H_{11}O$, 36), 158 (M^+ − $C_6H_{10}O$, 36), 157 (M^+ − $C_6H_{11}O$, 100), 146 ([M + 2]$^+$ − $C_7H_{12}O$, 24), 144 (M^+ − $C_7H_{12}O$, 73), 94 (16), 67 (11). Anal. ($C_{11}H_{17}ClN_4O$) C, H, N.

(±)-*cis*-6-Chloro-9-[2-(hydroxymethyl)cyclopentyl]-9*H*-purine (6a). A mixture of **5a** (150 mg, 0.62 mmol), $CH(OEt)_3$ (3.4 mL, 0.02 mol) and 12 M HCl (0.04 mL) was stirred for 12 h at room temperature; the solvent was then evaporated under vacuum and the solid residue redissolved in THF (10 mL) and 0.5 M HCl (13 mL). After 2 h of stirring at room temperature, the mixture was neutralized with 0.5 M NaOH, the solvent evaporated under vacuum and the solid residue purified by FC with 98:2 CH_2Cl_2/CH_3OH as eluent, which produced pure **6a** (125 mg, 80%): mp 119−121 °C. ^{1}H NMR (Cl_3CD): 8.75 (s, 1H, H-2), 8.10 (s, 1H, H-8), 5.20 (m, 1H, CH−N), 3.45 (m, 2H, CH_2−O), 2.85 (m, 1H, OH aliphatic), 2.57−1.42 (m, 7H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (Cl_3CD): 152.6, 152.0, 151.8, 145.0, 131.8, 61.2, 58.4, 48.0, 31.3, 26.0, 22.7. IR: 3283, 2954, 1592, 1567, 1394, 1335, 1229, 632. MS *m*/*z* (%): 254 ([M + 2]$^+$, 6), 252 (M^+, 18), 181 (31), 157 ([M + 2]$^+$ − C_6H_9O, 35), 155 (M^+ − C_6H_9O, 100), 149 (17), 119 (15), 83 (29), 67 (14). Anal. ($C_{11}H_{13}ClN_4O$) C, H, N.

(±)-*cis*-6-Chloro-9-[2-(hydroxymethyl)cyclopentylmethyl]-9*H*-purine (6c). This compound was prepared from **5c** (500 mg, 1.95 mmol) in an analogous way to **6a**. The crude product was purified by FC with 99:1 CH_2Cl_2/CH_3OH as eluent, which gave pure **6c** (500 mg, 96%): mp 78−79 °C. ^{1}H NMR (Cl_3CD): 8.73 (s, 1H, H-2), 8.24 (s, 1H, H-8), 4.61 (dd, 1H, HC*H*-N, *J* = 13.85 and 6.15), 4.25 (dd, 1H, *H*CH−N, *J* = 13.85 and 9.50), 3.74 (d, 2H, CH_2−O, *J* = 6.55), 3.45 (t, 1H, OH, *J* = 6.55), 2.65 (m, 1H, CH−C−N), 2.29 and 1.92−1.39 (2m, 2+5H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (Cl_3CD): 152.4, 152.2, 151.3, 146.2, 131.9, 63.5, 45.8, 43.7, 42.7, 29.7, 28.3, 23.0. IR: 3287, 2930, 1595, 1560, 1401, 1334, 1185, 1058, 945. MS *m*/*z* (%): 268 ([M + 2]$^+$, 7), 266 (M^+, 21), 235 (M^+ − CH_3, 14), 170 ([M + 2]$^+$ − $C_6H_{10}O$, 17), 168 (M^+ − $C_6H_{10}O$, 48), 167 (M^+ − $C_6H_{11}O$, 41), 157 ([M + 2]$^+$ − $C_7H_{11}O$, 34), 155 (M^+ − $C_7H_{11}O$, 100), 81 (17), 79 (19), 67 (12). Anal. ($C_{12}H_{15}ClN_4O$) C, H, N.

(±)-*trans*-6-Chloro-9-[2-(hydroxymethyl)cyclopentylmethyl]-9*H*-purine (6d). This compound was prepared from **5d** (500 mg, 1.95 mmol) in an analogous way to **6a**. The crude product was purified by FC with 98:2 CH_2Cl_2/CH_3OH as eluant, which gave pure **6d** (510 mg, 98%): mp 99−101 °C. ^{1}H NMR (DMSO-d_6): 8.76 and 8.72 (2s, 2H, H-2 and H-8), 4.35 (dd, 1H, HC*H*-N, *J* = 13.70 and 7.00), 4.18 (dd, 1H, *H*CH−N, *J* = 13.70 and 8.20), 3.17 (m, 2H, CH_2−O), 2.25 (m, 1H, CH−C−N), 1.80−1.25 (m, 7H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 152.5, 151.8, 149.2, 148.1, 131.0, 64.5, 48.5, 45.5, 42.5, 30.5, 29.1, 24.1. IR: 3320, 2945, 1601, 1568, 1402, 1341, 939, 646. MS *m*/*z* (%): 268 ([M + 2]$^+$, 8), 266 (M^+, 23), 248 (M^+ − H_2O, 23), 170 ([M + 2]$^+$ − $C_6H_{10}O$, 21), 169 ([M + 2]$^+$ − $C_6H_{11}O$, 26), 168 (M^+ − $C_6H_{10}O$, 57), 167 (M^+ − $C_6H_{11}O$, 55), 157 ([M + 2]$^+$ − $C_7H_{11}O$, 38), 155 (M^+ − $C_7H_{11}O$, 100), 79 (21), 67 (9). Anal. ($C_{12}H_{15}ClN_4O$) C, H, N.

(±)-*cis*-9-[2-(Hydroxymethyl)cyclopentyl]adenine (7a). Gaseous ammonia was bubbled for 1 h through a solution of **6a** (100 mg, 0.40 mmol) in methanol (20 mL) in a steel reactor at −80 °C. The reactor was closed and heated for 20 h in an oven at 60 °C, the solvent was evaporated under vacuum, and the solid residue purified by FC with 98:2 Cl_2CH_2/CH_3OH as eluant, which gave pure **7a** (90 mg, 98%): mp 208−209 °C. ^{1}H NMR (DMSO-d_6): 8.11 and 8.07 (2s, 2H, H-2 and H-8), 7.18 (s, 2H, NH_2), 4.95 (m, 1H, CH−N, *J* = 7.10), 4.40 (t, 1H, OH aliphatic, *J* = 5.05), 2.94 (m, 2H, CH_2−O), 2.38−1.59 (m, 7H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 156.3, 152.5, 150.3, 140.5, 118.9, 60.8, 56.5, 45.7, 30.7, 27.5, 22.6. IR: 3109, 2966, 1682, 1614, 1302, 1038, 691. MS *m*/*z* (%): 233 (M^+, 34), 216 (M^+ − NH_3, 43), 203 (M^+ − CH_2, 14), 162 (73), 136 (M^+ − C_6H_9O, 78), 135 (M^+ − $C_6H_{10}O$, 100), 108 (33), 81 (10), 67 (13). Anal. ($C_{11}H_{15}N_5O$) C, H, N.

(±)-*cis*-9-[2-(Hydroxymethyl)cyclopentylmethyl]-adenine (7c). This compound was prepared from **6c** (200 mg, 0.75 mmol) in an analogous way to **7a**. The crude product was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **7c** (174 mg, 94%): mp 184−185 °C. ^{1}H NMR (DMSO-d_6): 8.13 and 8.11 (2s, 2H, H-2 and H-8), 7.13 (bs, 2H, NH_2), 4.56 (m, 1H, OH aliphatic), 4.28 (dd, 1H, HC*H*-N, *J* = 13.55 and 5.55), 4.03 (dd, 1H, *H*CH−N, *J* = 13.55 and 10.65), 3.52−3.37 (m, 2H, CH_2−O), 2.55 (m, 1H, CH−C−N), 2.10 (m, 1H, CH−C−O), 1.76−1.22 (2m, 2+4H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 156.3, 152.6, 150.0, 141.3, 119.0, 61.4, 44.0, 43.7, 41.8, 29.1, 28.0, 22.6. IR: 3128, 1670, 1605, 1418, 1305, 1240, 644. MS *m*/*z* (%): 247 (M^+, 30), 230 (M^+ − NH_3, 20), 216 (M^+ − CH_3O, 24), 176 (27), 149 (M^+ − $C_6H_{10}O$, 86), 148 (M^+ − $C_6H_{11}O$, 91), 136 (M^+ − $C_7H_{11}O$, 64), 135 (M^+ − $C_7H_{12}O$, 100), 108 (28), 79 (11), 67 (12). Anal. ($C_{12}H_{17}N_5O$) C, H, N.

(±)-*trans*-9-[2-(Hydroxymethyl)cyclopentylmethyl]-adenine (7d). This compound was prepared from **6d** (100 mg, 0.37 mmol) in an analogous form to **7a**. The crude product was purified by FC with 98:2 CH_2Cl_2/CH_3OH as eluent, which gave pure **7d** (90 mg, 97%): mp 199−200 °C. ^{1}H NMR (DMSO-d_6): 8.13 and 8.11 (2s, 2H, H-2 and H-8), 7.17 (s, 2H, NH_2), 4.50 (t, 1H, OH, *J* = 5.05), 4.17 (dd, 1H, HC*H*-N, *J* = 13.60 and 6.65), 4.01 (dd, 1H, *H*CH−N, *J* = 13.60 and 8.30), 3.17 (m, 2H, CH_2−O), 2.17 (m, 1H, CH−C−N), 1.78−1.24 (m, 2+3+2H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 156.3, 152.7, 150.0, 141.4, 118.9, 64.5, 47.5, 45.5, 42.5, 30.4, 29.2, 24.1. IR: 3275, 2936, 1697, 1616, 1574, 1439, 1348, 1302, 1069, 652. MS *m*/*z* (%): 247 (M^+, 26), 230 (M^+ − NH_3, 19), 149 (M^+ − $C_6H_{10}O$, 100), 148 (M^+ − $C_6H_{11}O$, 97), 136 (M^+ − $C_7H_{12}O$, 52), 135 (M^+ − C_7H_9O, 54), 108 (18), 77 (11), 67 (10). Anal. ($C_{12}H_{17}N_5O$) C, H, N.

(±)-*cis*-9-[2-(Hydroxymethyl)cyclopentyl]hypoxanthine (8a). A mixture of **6a** (100 mg, 0.40 mmol) and 0.5 M NaOH (5 mL) was refluxed for 5 h. The solvent was then evaporated under vacuum and the solid residue purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **8a** (87 mg, 97%): mp 265−266 °C. ^{1}H NMR (DMSO-d_6): 12.19 (bs, 1H, OH aromatic), 8.05 and 8.01 (2s, 2H, H-2 and H-8), 4.94 (m, 1H, CH−N, *J* = 7.25), 4.33 (m, 1H, OH aliphatic), 2.95 (m, 2H, CH_2−O), 2.38−1.58 (m, 7H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 156.8, 148.8, 145.2, 139.6, 123.7, 60.5, 56.5, 45.0, 30.6, 27.3, 22.3. IR: 3392, 2874, 1716, 1654, 1597, 1550, 1418, 1136, 602. MS *m*/*z* (%): 234 (M^+, 32), 204 (M^+ − CH_2O, 8), 205 (M^+ − CHO, 6), 164 (29), 163 (30), 137 (M^+ − C_6H_9O, 100), 136 (M^+ − $C_6H_{10}O$, 70), 109 (26), 81 (15), 67 (16). Anal. ($C_{11}H_{14}N_4O_2$) C, H, N.

(±)-*cis*-9-[2-(Hydroxymethyl)cyclopentylmethyl]-hypoxanthine (8c). This compound was prepared from **6c** (200 mg, 0.75 mmol) in an analogous way to **8a**. The crude product was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **8c** (165 mg, 89%): mp 296−298 °C. ^{1}H NMR (DMSO-d_6): 12.25 (bs, 1H, OH aromatic), 8.10 and 8.01 (2s, 2H, H-2 and H-8), 4.59 (m, 1H, OH aliphatic), 4.29 (dd, 1H, HC*H*-N, *J* = 13.55 and 5.45), 4.03 (dd, *H*CH−N, *J* = 13.55 and 10.85), 3.42 (m, 2H, CH_2−O), 2.52 (m, 2H, CH−C−N), 2.10 (m, 1H, CH−C−O), 1.71−1.21 (m, 6H, $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 156.5, 148.3, 145.2, 140.3, 123.7, 60.9, 43.9, 43.1, 41.5, 28.5, 27.4, 22.1. IR: 3406, 2947, 2864, 1673, 1592, 1413, 1225, 757. MS *m*/*z* (%): 248 (M^+, 19), 219 (21), 178 (17), 150 (M^+ − $C_6H_{10}O$, 36), 149 (M^+ − C_6H_9O, 36), 137 (M^+ − $C_7H_{11}O$, 100), 136 (M^+ − $C_7H_{12}O$, 50), 109 (27), 81 (14), 67 (19). Anal. ($C_{12}H_{16}N_4O_2$) C, H, N.

(±)-*trans*-9-[2-(Hydroxymethyl)cyclopentylmethyl]-hypoxanthine (8d). This compound was prepared from **6d** (200 mg, 0.75 mmol) in an analogous way to **8a**. The crude product was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **8d** (180 mg, 97%): mp 237−238 °C. ^{1}H NMR (DMSO-d_6): 11.82 (bs, 1H, OH aromatic), 8.09 and 8.02 (2s, 2H, H-2 and H-8), 4.17 (dd, 1H, HC*H*-N, *J* = 13.60 and 6.80), 4.01 (dd, 1H, *H*CH−N, *J* = 13.60 and 8.35), 3.16 (m, 2H, CH_2−O), 2.17 (m, 1H, CH−C−N), 1.77−1.22 (m, 7H, CH−C−O + $(CH_2)_3$). ^{13}C NMR (DMSO-d_6): 157.1, 148.9, 145.7, 140.8, 124.1, 64.5, 47.9, 45.5, 42.7, 30.4, 29.2, 24.1. IR: 3408, 2862, 1680, 1591, 1435, 1221, 1126, 1067. MS *m*/*z* (%): 248 (M^+, 23), 219 (M^+ − CHO, 9), 150 (M^+ − $C_6H_{10}O$, 39), 149 (M^+ − $C_6H_{10}O$, 61), 137 (M^+ − $C_7H_{11}O$, 100), 109 (18), 67 (14). Anal. ($C_{12}H_{16}N_4O_2$) C, H, N.

Computational Approach. All computations were carried out on a PC Pentium III 240 MHz. The TOSS-MODE for Windows package developed in our laboratory was used to compute the molecular descriptors (spectral moments) for the data sets of compounds. The linear discriminant analysis was performed by using the STATISTICA for Windows package.[28] Default parameters of this program were always used.

Antitumor Activity Test. All assays were performed in flat-bottomed 96-well microtiter plates. To each well were added 5×10^4 L1210 cells or 7.5×10^4 CEM or Molt4/C8 cells and a certain amount of the test compound. The cells were allowed to proliferate for 48 h (L1210) or 72 h (CEM and Molt4/C8) at 37 °C in a humidified CO_2-controlled atmosphere. At the end of the incubation period, the cells were counted in a coulter counter (Coulter Electronics Ltd., Harpenden, Herts, England). The IC_{50} was defined as the concentration of compound that reduced the number of living cells by 50%.[43]

Supporting Information Available: The complete list of compounds used in training and prediction sets, as well as their a posteriori classification according to model 2, and elemental analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kubinyi, H. Chance favors the prepared mind − from serendipity to rational drug design. *J. Recept. Signal Transduct. Res*. **1999**, *19*, 15−39.

(2) Lien, E. J.; Lien. L. L. L. New drug design, discovery and development: retrospective and perspective views. *Chin. Pharm. J. (Taipei)* **1998**, *50*, 249−256.

(3) Lunney, E. A. Structure-based drug design begins a new era. *Med. Chem. Res*. **1998**, *8*, 352−361.

(4) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening-An Overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(5) Drie, J. H. V.; Lajiness, M. S. Approaches to virtual library design. *Drug Discovery Today* **1998**, *3*, 274−283.

(6) Julián-Ortiz, J. V.; Gálvez, J.; Muñoz-Collado, C.; García-Domenech, R.; Gimeno-Cardona, C. Virtual Combinatorial Synthesis and Computational Screening of NewPotential Anti-Herpes Compounds. *J. Med. Chem*. **1999**, *42*, 3308−3314.

(7) Denny, W. A. The role of medicinal chemistry in the discovery of DNA-active anticancer drugs: from random searching, through lead development, to de novo design. In *The Search for Anticancer Drugs*; Waring, M. J., Ponder, B. A. J., Eds.; Kluwer Academic Publishers: Dordrecht, 1992; Chapter 2.

(8) Menta, E.; Palumbo, M. Antineoplastic agents 1998. *Expert Opin. Ther. Pat*. **1998**, *8*, 1627−1672.

(9) Balasubramanian, B. N.; Kadow, J. F.; Kramer, R. A.; Vyas, D. M. Recent developments in cancer cytotoxics. *Annu. Rep. Med. Chem*. **1998**, *33*, 151−162.

(10) Ferrante, K.; Winograd, B.; Canetta, R. Promising new developments in cancer chemotherapy. *Cancer Chemother. Pharmacol*. **1999**, *43* (Suppl.), S61−S83.

(11) Ishioka, C.; Kato, S.; Kanamura, R. Current strategies on discovery of anticancer drugs in the United States of America. *Saishin Igaku* **1998**, *53*, 1940−1945.

(12) Buolamwini, J. K. Novel anticancer drug discovery. *Curr. Opin. Chem. Biol*. **1999**, *3*, 500−509.

(13) Estrada, E.; Peña, A.; García-Domenech, R. Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. *J. Comput.-Aided Mol. Des*. **1998**, *2*, 583−595.

(14) 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.

(15) 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.

(16) 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.

(17) Estrada, E. Modelling the diamagnetic susceptibilities of organic compounds by a substructural graph theoretical approach. *J. Chem. Soc., Faraday Trans*. **1998**, *94*, 1407−1411.

(18) Estrada, E.; Gutierrez, Y. Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *J. Chromatogr. A* **1999**, *858*, 187−199.

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

(20) Ferguson, L. N. *The Modern Structural Theory of Organic Chemstry*; Prentice Hall: Englewood Cliffs, NJ, 1963; pp 200−203.

(21) 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.

(22) Albert, A. *Selective Toxicity. The Physico-Chemical Basis of Therapy*; Chapman and Hall: London, 1984.

(23) Willet, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 983−996 and references therein.

(24) Estrada, E. Characterization of 3D molecular structure. *Chem. Phys. Lett*. **2000**, *319*, 713−718.

(25) Estrada, E. Generalized spectral moments of the iterated line graph sequence. A novel approach to QSPR studies. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 90−95.

(26) Randic, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 672−687.

(27) Negwer, M. *Organic-Chemical Drugs and their Synonyms*; Akademie-Verlag: Berlin, 1987.

(28) STATISTICA, version 4.13; Statsoft Inc., 1993.

(29) Craig, P. N. Substructural Analysis and Compound Selection. In *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds. Vol. 4. Quantitative Drug Design*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990; pp 645−666.

(30) Chu, K. C.; Feldman, R. J.; Shapiro, M. B.; Haza, G. F. pattern recognition and structure−activity relationship studies. Computer-assisted prediction of antitumor activity in structurally diverse drugs in an experimental mouse brain tumor system. *J. Med. Chem*. **1975**, *18*, 539−545.

(31) Leonard, D. M. Ras Farnesyltransferase: A New Therapeutic Target. *J. Med. Chem*. **1997**, *40*, 2971−2990.

(32) Gibbs, J. B.; Graham, S. L.; Hartmen, G. D.; Koblan, K. S.; Kohl, N. E.; Omer, C. A.; Oliff, A. Franesyltransferase inhibitors verus Ras inhibitors. *Curr. Opin. Biol. Chem*. **1997**, *1*, 197−203.

(33) Singh, S. B.; Zink, D. L.; Willians, M.; Polishook, J. D.; Sanchez, M.; Silverman, K. C.; Lingham, R. B. Kampanols: novel Ras farnesyl-protein transferase from Stachybotrys kampalensis. *Bioog. Med. Chem. Lett*. **1998**, *8*, 2071−2076.

(34) Kaminski, J. J.; Rane, D. F.; Snow, M. E.; Weber, L.; Rothofsky, M. L.; Anderson, S. D.; Lin, S. L. Identification of Novel Farnesyl Protein Transferase Inhibitors Using Three-Dimensional Database Searching Methods. *J. Med. Chem*. **1997**, *40*, 4103−4112.

(35) Teijeira, M. Nucleosidos carbocíclicos 1,2-disubstituidos. Ph.D. Thesis, Universidad de Santiago de Compostela, Spain, 1996.

(36) Shealy, Y. F.; O'Dell, C. A. Synthesis of the Carbocyclic Analogues of Uracil Nucleosides. *J. Heterocycl. Chem*. **1976**, *13*, 1015−1020.

(37) Hronowski, L. J. J.; Szarek, W. A. Regioespecific Synthesis of Cyclopentane Analogues of (2′- and 3′-deoxy-treo-pentofuranosyl)-uracil and -2-thiouracil Nucleosides. *Can. J. Chem*. **1985**, *63*, 2787−2797.

(38) Herdewijn, P.; De Clercq, E.; Balzarini, J.; Vanderhaeghe, H. Synthesis and Antiviral Activity of the Carbocyclic Analogues of (E)-5-(2-halovinyl)-2′-deoxyuridines and (E)-5-(2-halovinyl)-2′-deoxycytidines. *J. Med. Chem*. **1985**, *28*, 550−555.

(39) Katagiri, N.; Nomura, M.; Sano. 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]-9H-adenine**.** *J. Med. Chem*. **1992**, *35*, 1882−1886.

(40) Vince, R.; Hua, M. Synthesis and Anti-HIV Activity of Carbocyclic 2′,3′-Didehydro-2′,3′-dideoxy 2,6-Disubstituted Purine Nucleosides. *J. Med. Chem*. **1990**, *33*, 17−21.

(41) Santana, L.; Teijeira, M.; Uriarte, E. A Sligtly Shorter Route to Carbocyclic Nucleosides. Synthesis of (±)-*trans*-1-[2-(Hydroxymethyl)cyclopentylmethyl]uracil. *J. Heterocycl. Chem*. **1999**, *36*, 293−295.

(42) Santana, L.; Teijeira, M.; Uriarte, E.; Terán, C.; Castellato, U.; Grazini, R. Synthesis of *cis*-1-[(2-hydroxymethyl)cyclopentyl]-uridine and determination of its conformation by X-ray crystallography and AM1 theoretical calculations. *Nucleosides Nucleotides* **1996**, *15*, 1179−1187.

(43) De Clercq, E. In *In vivo and ex vivo test systems to rationalize drug design and delivery*; Crommelin, D., Couvreur, P., Duchene, D., Eds.; Editions de Santé: Paris, France, 1994; pp 108−125.

JM991172D