Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 14 (2006) 1095–1107
Bioorganic & Medicinal Chemistry

Stochastic entropy QSAR for the in silico discovery of anticancer compounds: Prediction, synthesis, and in vitro assay of new purine carbanucleosides

Humberto González-Díaz,[a,b,*] Dolores Viña,[b,c] Lourdes Santana,[b] Erik de Clercq[d] and Eugenio Uriarte[b,*]

[a]*Department of Drug Design, Chemical Bioactives Center, Central University of "Las Villas," Villa Clara 54830, Cuba*
[b]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Santiago 15782, Spain*
[c]*Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Salamanca, Salamanca 37008, Spain*
[d]*Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven B-3000, Belgium*

Received 19 May 2005; revised 12 September 2005; accepted 13 September 2005
Available online 25 October 2005

Abstract—A Markov model based QSAR is introduced for the rational selection of anticancer compounds. The model discriminates 90.3% of 226 structurally heterogeneous anticancer/non-anticancer compounds in training series. External validation series were used to validate the model; the 91.8% containing 85 compounds, not considered to fit the model, were correctly classified. The model developed is afterwards used in a simulation of a virtual search for anticancer compounds never considered either in training or in predicting series. The 87.7% of the 213 anticancer compounds used in this simulated search were correctly classified. The model also shows high values for specificity (0.89), sensitivity (0.91), and Mathews correlation coefficient (0.79). In addition, the present model compares better-to-similar with respect to other four models elsewhere reported if one takes into consideration 26 comparison parameters. Finally, we exemplify the use of the model in practice with the design of a new series of carbanucleosides. The compounds evaluated with the model were synthesized and experimentally assayed for their antitumor effects on the proliferation of murine leukemia cells (L1210/0) and human T-lymphocyte cells (CEM/0 and Molt4/C8). The more interesting activity was detected for the compound **5a** with a predicted probability of 80.2% and IC_{50} = 27.0, 27.2, and 29.4 μM, respectively, against the above-mentioned cellular lines. These values are comparable to those for the control compound Ara-A.

1. Introduction

Every year millions of chemicals are added to the Chemicals Abstract Service registry. Considering that the measurements of biological activities and toxicological parameters, either in vivo or in vitro, are in general extremely time-consuming and expensive, it is imperative to develop novel alternative techniques.[1] Thence, a major impetus to the development of replacement alternatives is provided by the speed with which the pharmaceutical industry can now produce new drug candidates. This, in turn, leads to a requirement for the development of methods to complement the high throughput screening technique to assess efficacy and potential toxicity. The new methods must be adaptable to the simultaneous and rapid testing of many thousands of compounds.[2]

Pharmaceutical industry and medicinal chemistry research groups are nowadays under particularly increasing pressure to discover new anticancer drug leads in a faster and more efficient way than in the past.[3–6] In this sense, computer-aided rational drug design strategies like quantitative-structure–activity relationships (QSARs) or docking approaches continuously attract the attention of medicinal chemists as one of the tools that may be used to find a promising solution to this problem.[7–12] In this connection, more than 1500 different molecular descriptors with potential applications in general to drug design and specifically anticancer drug discovery have appeared in the literature. Among the main classes of molecular descriptors topologic indices,

Keywords: QSAR; Anti-cancer compounds; Markov model; Nucleoside; Leukemia.
* Corresponding authors. Tel.: +34 981 563 100; fax: +34 981 594912; e-mail addresses: gonzalezdiaz@yahoo.es; qohumbe@usc.es

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.09.039

information theory related descriptors, graph-theoretical invariants, and quantities derived by quantum or molecular mechanics stand out.[13]

Certainly, QSARs based on a series of homologue compounds, which have possibly the same mechanism of action, are easier to derive but they are however limited only to the discovery of a narrow range of anticancer compounds. Conversely, general-purpose QSAR models describing the anticancer activity of large and heterogeneous series of compounds are in principle applicable for the discovery of anticancer compounds with different structural patterns. Nevertheless, there have been reported up-to-date just a few of general-purpose QSAR models.[14–16] In addition, there are several molecular descriptors, which have still not been tested for their skills to explain anticancer activity in general-purpose QSAR models. Consequently, the necessity of exploring the potentialities of other molecular descriptors to seek novel QSARs as a complement for the few existing general-purpose QSAR models in the virtual search of candidates to be synthesized and experimentally assayed as anticancer drugs arises.[17–20]

In this connection, our research group has well documented the use of a Markov chain (MC) model, to derive new molecular descriptors for drug design. MC models constitute one of the most used stochastic theories to encode molecular structure.[21] The approach referred to above was termed as Markovian-chemicals-in silico-design (MARCH-INSIDE). Some of the molecular descriptors derived with this approach have shown to be very useful in the design of novel antimicrobial drugs.[22] There have also been reported interesting applications in mathematical biology.[23] The reinterpretation of the method in terms of entropy like molecular descriptors allowed toxicology,[24] bioinformatics,[25] proteomics,[26] and virology[27] applications. In any case, of the major interest for the present work was an early report of a general-purpose 2D-QSAR model for the selection of anticancer compounds using the MARCH-INSIDE molecular descriptors.[28] Unfortunately, this model (because of its 2D nature) is unable to explain the anticancer activity of several compounds with 3D structural features such as chirality and *Z–E* isomerism. However, our group has after that reported the 3D version of MARCH-INSIDE.[29,30] This fact encourages us to test the potentialities of this new 3D-QSAR approach in the study of highly 3D-structure-dependent pharmacological activities like anticancer action.

2. Results and discussion

2.1. QSAR for anticancer discovery

QSAR approaches based on different molecular descriptors are of major importance for bioorganic and medicinal chemistry nowadays.[31] Particularly, QSAR methodologies based on stochastic approaches to generate molecular descriptors have found many applications in this field recently.[32–37] Specifically, the molecular descriptors used in the present work are the so-called stochastic entropies Θ_k, which are Shannon entropies describing the connectivity and the distribution of electrons for each atom in the molecule.[38,39]

The present study attempts to develop a simple linear QSAR to discriminate anticancer compounds from non-active ones. With this purpose we collected from recognized databases a series of 157 anticancer compounds and 153 having biological activities other than anticancer.[40,41] This original series was divided at random into training and predicting series. Training series was composed of 108 anticancer compounds and 118 non-anticancer compounds. This series was used to fit the QSAR by means of linear discriminant analysis (LDA)[42,43]:

$$\mathrm{ACAN} = 7.859 \times {}^{3}O(\Theta_5) + 9.791 \times {}^{2}O(\Theta_1) + 1.368, \quad (1)$$
$$N = 226, \quad \lambda = 0.45, \quad F = 135.0, \quad p < 0.001.$$

The model has shown to be statistically significant at the *p*-level ($p < 0.001$) with Wilk's statistics $\lambda = 0.45$ and Fisher ratio $F = 135.0$. In Eq. 1, the symbol ${}^{\mathrm{I}}O(\Theta_k)$ indicates that the variable (Θ_k) is selected with order of importance (I) after previous orthogonalization (O) by Randić's approach. Orthogonalization is used in order to avoid inter-variable co-linearity.[44–47] We have also inspected the percentage of good classification.

The model correctly classifies 90.3% of compounds in training series. Specifically, the model correctly classifies 89.8% of active compounds (97 out of 108). The names and posterior probabilities for these compounds are depicted in Table 1. At the same time, the model correctly classifies 90.7% of non-active ones (107 out of 118) (see Table 2). The appearance of chance correlation was prevented by exploring only the first six different Θ_k including Θ_0–Θ_5. In this case, six variables $\ll 112$ cases, it has been previously shown that the number of variables to be explored must be less than $(N/2) - 1$ the number of observations (N).[42]

The total predictability was 91.8%; in detail, the model correctly classifies 91.8% (45 out of 49) anticancer compounds and 91.7% of non-active compounds (33 out of 36). Results for each one of these compounds are depicted in Table 3.

2.2. Comparison with other QSAR studies and virtual screening

The selection of large and structurally heterogeneous series of compounds in order to develop QSAR studies could be the source of discussion. However, it has been largely demonstrated that the use of very heterogeneous series of organic compounds may yield potent QSARs, even for very specific properties.[33] In any case, we are going to focus only on anticancer activity. There have been reported, until our concern, three QSAR models based on topologic molecular descriptors for the discovery of anticancer compounds using heterogeneous series of chemicals having different mechanisms of action: (a) the 2D-MARCH-INSIDE methodology,[28] (b) the TOPS-MODE approach,[48] and (c) the connectivity

Table 1. Name and posterior probabilities for anticancer chemicals in training series

Training series anticancer compounds	
Name	*p*
Lost	0.59
Cystogon	0.58
Tedegyl	0.53
ACIA	0.39
Elmustine	0.84
Mitoguazone	0.81
DON	0.29
Athoxen	0.65
Mitolactol	0.89
Trichlormetine	0.93
Aphoxide	0.80
Deltespamide	0.84
C61	0.94
Embichin 11	0.90
Neo-SHP	0.89
Busulfan	0.91
Amebisan	0.91
NSC-12272	0.38
Embichin 7	0.95
Aldophosphamide	0.96
Oxisuran	0.25
Uramustine	0.95
Azacitidine	0.88
DMDAI	0.94
OPSPA	0.93
Spiroplatin	0.98
1954 C.B.	0.53
Betamin	0.38
Floxuridine	0.88
Aucitabine	0.21
Cytarazid	0.95
Thyminalkylamine	0.98
BR2 51308	0.97
Dopastin	0.98
Echinosporin	0.81
Acetoxycycloheximide	0.99
Buthiopurine	0.46
Aniline mustard	0.54
Novembitol	0.56
Ederpin	0.99
GYKI 13324	1.00
Thiohexadepa	0.98
Ritrosulfan	0.99
Metropine	0.38
Neplanocin	0.97
3-Deazaguanosine	0.97
Raxosan	0.89
Meturedepa	0.97
Triaziquone	0.63
IOB 177	0.33
Sangivamycin	0.98
Pumitepa	0.98
Thiodirin	0.99
Dithizon	0.64
Mitotenamide	0.67
Triciribine phosphate	0.98
Clofenotane	0.44
Frolon	0.99
Simtrazene	0.76
Nitrocafan	0.66
Ocaphane	0.77
Spiromustine	1.00
Lapachol	0.53
Rifazepam	0.99
Carboquone	0.83
Hydroxycycloheximide	0.98
Ptilocaoline	0.97
Febrifugine	0.75
Inproquone	0.83
Spirazidin	1.00
Piritrexim	1.00
Azotomycin	0.99
U-29409	0.63
Aminoanfol	0.99
CGP 15720	0.99
Diethylstylbestrol diSO_4	0.76
Quinocarcin	0.76
Prosfidium	1.00
Pafencil	1.00
Testolactone	0.86
Camptothecin	0.71
Acodazole	0.00
Bremfol	1.00
Ninofterin	0.99
Zimet 54/79	1.00
Benfluron	0.80
Methasquin	0.83
Fenastezin	0.95
BAF	0.92
Ketotrexate	1.00
NSC-83265	0.93
GEA 29	1.00
Eupochlorin acetate	1.00
Chlorbutinpenicillin	0.97
Homocoralyne	1.00
Isopropylcad	1.00
Drostanolone	1.00
Sibiromycin	1.00
Osaiin	0.81
Astiron	0.96
Dauronobicin	1.00
Tetracycline	0.95
Menogaril	1.00
Pactamycin	1.00
Teralpherin	1.00
m-Embitol	0.72
o-Embitol	0.76
p-Embitol	0.85

Table 1 (*continued*)

p is the posterior probability by which the chemical is non-anticancer. The *p*-level threshold limit is p 0.05, thence a chemical with $0.55 < p < 0.45$ may be considered as non-classified.

approach.[49] In Table 4, we compile more than 20 parameters for comparison with respect to 3D-MACH-INSIDE. All these parameters may be grouped on complexity, training, validation features, and other topics:

(i) *In terms of complexity.* The 3D-MARCH-INSIDE approach reported here (model 1 in Table 4) is the less complicated one with only two descriptors. The 2D-MARCH-INSIDE (model 2) is notably more complicated with six variables. The 2D-MARCH-INSIDE is, in any case, less complicated than model 3 (TOPS-MODE approach), which is the more complicated with 10 variables in the model, with two variables the product of

Table 2. Name and posterior probabilities of non-anticancer chemicals in training series

Training series non-anticancer compounds	
Name	*p*
Selectan	0.95
Metofurone	0.95
Allisan	0.94
AB41	0.89
F-28	0.83
S131	0.95
Bromocyl	0.95
Apasicin (a)	0.88
Methamilane	1.00
Chlorphenacemide	0.72
PF-527	0.85
Metacetanilidum	1.00
Strinoline	0.84
Lofemizole	0.83
BW 775C	0.81
4-Aminophenazone	0.44
Sulfafurazole	0.54
Alfazole	0.51
Fludorex	0.57
Diclonixin	0.68
Brofezil	0.58
Diphenizin	0.56
Metazide	0.89
Orthovanizide	0.50
Dibatod	0.78
Ceftriaxone	1.00
Brindoxime	0.18
Metrizamide	0.38
Basedol	0.99
Ferban	0.96
Phenylbutazone	0.92
Serotonin	0.73
Trichlormethiazide	0.86
Asa	0.85
Anot	0.81
Ioxynil	0.95
Linuron	0.77
Maleic hidrazide	0.96
MCP	0.88
Genite	0.74
Monuron	0.73
Morestan	0.90
Ovex	0.78
Chloral Hidrate	0.95
Picloram	0.95
Chlorothiazide	0.83
Sirmate	0.77
Allobarbital	0.40
Sweep	0.88
Tetrachlorothiophene	0.99
Tetradifon	0.67
Acrolein	0.99
Barban	0.69
Dichlorphenamine	0.91
Bromoxynil	0.96
C-56	0.95
C-57	0.95
Hydrochlorothiazide	0.75
Carbophenithion	0.94
Chloranil	0.98
Meparfynol	0.97
Acetanilide	0.80
Merbromin	0.33
Chlorthion	0.83
CIPC	0.58
Methyclothiazide	0.80
Methyl salicylate	0.91
Nalidixic acid	0.70
2.4-D	0.87
Nitrofurazone	0.96
Dacthal	0.90
Dalapon	0.95
Phenindione	0.70
Polythiazide	0.69
Guthion	0.60
Folpet	0.93
Sulfaguanidine	0.57
Dichobenil	0.96
Sulfanilamide	0.78
p-Dichlorobenzene	0.97
Trichloroethanol	1.00
Amiben	0.90
Zoxazolamine	0.96
Amitrole	0.99
Diuron	0.72
Dyrene	0.80
Fenac	0.89
Ac.Aminosalisylic	0.94
Ac.Meclofenamic	0.60
Aciclovir	0.92
Alupurinol	0.97
Benzene	0.95
Chlorothiazide	0.92
Chloramphenicol (AM)	0.49
Diazoxide	0.90
Estibufen	0.74
Flunitrazepan	0.63
Furosemide	0.72
Halotane	0.99
Benzene hexachloride	0.98
Hidrochlorothiazide	0.92
Isoniazida	0.89
Levodopa	0.79
Niridazol	0.93
Nitrofurantoin	0.81
Nitroglicerin	0.81
Nitroprusiate	0.86
Pirazinamide	0.96
Piroxican	0.50
Prontosil	0.44
Resorcinol	0.96
Sulfasalazine	0.35
Tetranitrate	0.46
Thiouracil	0.98
Tolrestat	0.59
Toluene	0.88
Zopicolone	0.21
Dichloran	0.94

Table 2 (*continued*)

p is the posterior probability by which the chemical is non-anticancer.

two different molecular descriptors. The model 4 (connectivity) is the second less complicated with only two variables but with one of them the sum of two different connectivity indices.

(ii) *In connection to training.* The 4 models were fitted with LDA and are statistically ($p < 0.00$) significant but the connectivity model effects more inefficient separation of anticancer/non-anticancer drugs according to

Table 3. Name and posterior probabilities of compounds in predicting series

Name	*p*
Anticancer compounds predicting series	
Bleuidon	0.53
Azaserine	0.31
Carmustine	0.80
Fluorembichin	0.71
Antimil	0.76
Chlormetine oxide	0.73
Pyramizid	0.70
Sparfosicum acid	0.70
Biocarbazine	0.81
CB1639	0.57
Caracemide	0.54
AI-1902	0.91
Thomizine	0.77
2-FPTS-Zn	0.22
Bromacrylide	0.83
Alanine Mustard	0.88
Marcophan	0.93
Chlormethyl Sliantrane	0.86
AB 100	0.90
B-518-ASTA	0.97
A-4942	0.96
Aethimidinum	0.69
Citofur	0.75
Azapycil	0.24
Leucenol	0.67
NSC-95466	0.95
Duazomycin	0.79
G-Azauridine	0.83
Imidazol Mustard	0.97
Fosfemid	0.46
SA 291	0.85
Azatepa	0.87
Streptozocin	0.96
ODEPA	0.91
Ethoxysilatrane	0.93
Sufosfamide	0.99
Nimustine	0.98
Chlorethylaminouracil	0.98
Damuar	0.91
Cytarabine	0.61
Pirazofurin	0.95
Lumostine	0.99
Mannityl dimesilate	0.97
Improsulfan	0.98
Broxuridine	0.92
Doxifluridine	0.84
Fluorafur	0.90
Azathiopirine	0.75
Neptamustine	0.99
Non-anticancer compounds predicting series	
Nitrofurylen	0.99
Furazonal	0.95
Bromaspirin	0.87
Atrolactimide	0.65
Tiabendazole	0.85
Dichlone	0.95
Nicosalicylicum acid	0.76
Oxasepam	0.65
Embramine	0.97
Phenacemide	0.70
HCA	0.96
Kepone	0.76
Parathion	0.60
Silvex	0.92
2,4,5-T	0.89
Zoalene	0.89
Ethylbiscoumacetate	0.22
B-Aminosalicylic acid	0.94
Meparfynol carbamate	0.97
Methaqualone	0.43
Morphine	0.34
Pentachlorophenol	0.97
Phenylhidrazine	0.86
Dicamba	0.94
Dicryl	0.79
Dinitrocresol	0.90
Dyrene	0.80
Acetazolamida	0.97
Bromazepan	0.65
Diflunisal	0.82
Ganciclovir	0.75
Hydralazine	0.87
Nitrazepan	0.66
Benzene oxide	0.98
Quazepan	0.63
Vit. K	0.84

its highest $\lambda = 0.63$. It is also notable that the 3D-MARCH-INSIDE is the only one with no-collinearity. The model also shows high values for specificity (0.89), sensitivity (0.91), and Mathews correlation coefficient (0.79). We also depict in Figure 1 the receiving operating characteristic curve (ROC curve) for the present model. Notably, the area under the ROC curve is significantly higher than the area under the random classifier curve (main diagonal). It means that the present QSAR is not a random, but a statistically significant, classifier.

(iii) *With respect to validation.* The validation of the four models was carried out by external predicting series and additional virtual screening series. The 3D-MARCH-INSIDE approach has been validated with the largest number of corroborated anticancer compounds $N1 + N2 = 262$. A total of 213 of these compounds were used for a simulated virtual screening of anticancer drugs in which the model showed a success of 87.7%, see Table 5.

(iv) *Other topics.* In this section, we would like to highlight that the present model is the only one of the four that encodes a 3D structure and can be at the same time back-projected to obtain maps representing the contribution of each atom in the molecule to the biological activity. Both aspects: codification of chirality and other 3D features[50] as well as model back-projection[51] are of major importance for the applicability of QSAR in medicinal chemistry (discussed below).

2.3. Discovery of novel carbanucleoside anticancer compounds

As an illustrative example of the use of the model in practice, we predict the biological activity for different 1,2-disubstituted carbanucleosides. The compounds were evaluated by the model, synthesized, and then assayed for anticancer activity (see Table 6). The 226 com-

Table 4. Comparison with respect to other approaches

	Anticancer activity classification models[a]			
Complexity				
Models' features to be compared	1	2	3	4
Explored variables	10	10	15	31
Variables in the model	2	6	10	2
Descriptors in the model	2	6	10	3
Variable product	No	No	2	No
Variable sum	No	No	No	1
Training				
Technique[b]	LDA	LDA	LDA	LDA
Wilk's λ (U-statistics)	0.45	0.44	0.56	0.63
F	135.0	141.3	11.2	74.6
p-level	>0.00	>0.00	>0.00	>0.00
collinearity	No	High	High	Low
Accuracy (%)	90.3	90.5	88.3	NR
Families of drugs[c]	Het.	Het.	Het.	Het.
N	226	681	222	264
N anticancer	108	—	63	NR
Validation				
Validation method[d]	i and ii	i	i and ii	i
$N1$ anticancer	49	84	21	NR
Predictability (%)	91.67	84.5	91.4	NR
$N2$ anticancer	213	0	30	No
Predictability (%)	87.7	—	87.0	No
Families of drugs[b]	Het.	—	1	—
$N1 + N2$	262	84	51	—
N total anticancer	370	298	116	NR
N total	524	961	323	NR
Mean success (%)	91.0	88.3	89.9	81.2
Other topics				
3D-structure[e]	Yes	No	No	No
Back-projection[f]	Yes	Yes	Yes	No

[a] Model 1 is reported in this work, models 2, 3, and 4 were reported in references 28, 48, and 49, respectively.
[b] LDA refers to linear discriminant analysis.
[c] Only largely represented families were considered.
[d] Validation methods are: (i) external predicting series, (ii) virtual screening.
[e] Considers the capability of the method to encode together chirality, Z–E, and axial–equatorial isomerism.
[f] Considers the possibility for deriving a map with the calculated contribution of any atom in the molecule to the biological activity.

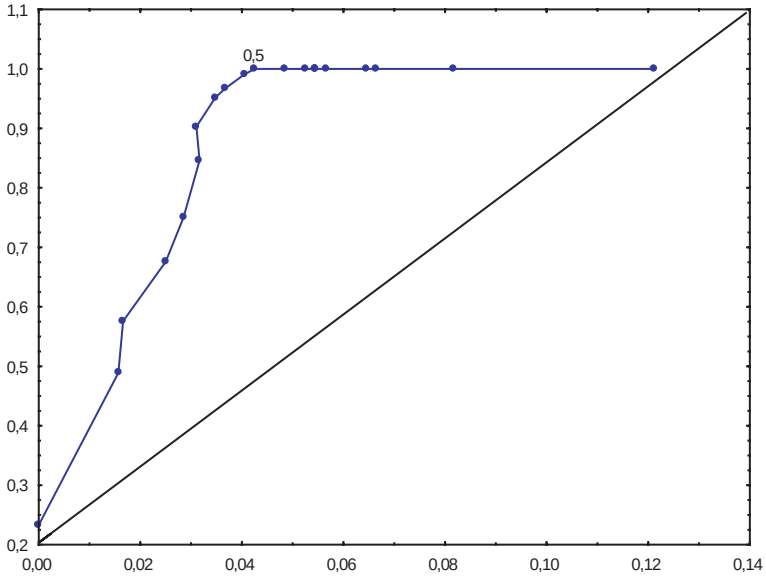


Figure 1. ROC curve for the QSAR model.

Table 5. Results for the virtual screening

Virtual screening of known anticancer compounds ($N = 213$, 87.7%)	
Name	*p*
2-AA	0.97
A-139	0.81
AB-182	0.96
AC 3579	0.89
Aceglatone	0.48
A-denopterin	1.00
Alalon	0.74
Alazopeptin	0.99
Albonoursin	0.99
Alkyrom	0.37
Altretamine	0.98
Ambunol	0.98
Amekrin	0.87
Ametantrone	0.93
Aminoalanfol	0.99
Aminochlorambucil	0.73
Aminofterin	0.99
Aminoglutethimide	0.92
Aminoterofetrin	1.00
Aminotreofol	0.99
AMSA	0.75
Amygdalin	1.00
Angustibalin	0.75
Antramycin	0.85
Ara-T	0.96
Arazide	0.96
Asalei	0.96
Asaline	0.97
Asamet	0.98
Asazol	0.87
Asdofan	0.96
Asperlin	0.84
AT-16	0.85
AT-346	0.24
Aziprin	1.00
Bactobolin	0.88
Benaxibine	0.96
Bendamustine	0.86
Bensolide	0.89
Benzodepa	0.95
Betacin	0.49
Benzodet	0.76
Benzotef	0.92
Bezatine tennazoate	0.97
Bimolane	1.00
Bisantrene	0.88
BM 123	0.98
Bromericum acid	0.46
Bufloracil	0.97
Bufumustine	0.76
Burseran	0.95
Butastezine	0.96
Butocin	0.82
Butoctamide	1.00
Butodicin	0.89
Butotricin	0.86
BW50-71	0.10
Calcil mefolinas	0.99
Calusterone	1.00
CAM	1.00
Carboelatin	0.29
Carmafur	0.44
CB 1837	0.09
CB-10252	0.87

Table 5 (*continued*)

Virtual screening of known anticancer compounds ($N = 213$, 87.7%)	
Name	*p*
Ceruicarcin	0.88
Chlorambucil	0.70
Chlorasquin	0.99
Chlorethylaminouracil	0.98
Chlornaphzine	0.63
Chlorozotocin	0.96
Chlorphenacyl	0.50
Chrysenex	0.59
Cianocycline	0.95
Citostal	0.68
Classen	0.04
Cleisthanthin	1.00
Cofplaton	0.90
Coralyne	1.00
CRC 7001	0.95
Crotefoxide	0.97
Cytarabine	0.61
Cytochalasin B	0.98
DABI	0.76
Damuar	0.91
Defosfamide	0.99
Demecolcine	0.74
Denofterin	1.00
Diazan	0.15
Diaziquone	1.00
Dichlorallyllawsone	0.39
Diethylstylbestrol	0.81
Diiodobezotef	0.91
Dimedazol	0.81
Dimetfolamine	0.99
Dimezol	0.87
Dinaphtimine	0.97
Diopterin	1.00
Dipin	0.99
Ditiomustine	1.00
Edikron	0.84
Elderfield pyrimidine	0.86
Ellipticine	0.72
Elliptinium acetate	0.79
Emcyt	0.92
Enpromate	0.91
Enterodiol	0.89
Enterolactone	0.94
Epiprofidine	1.00
Esorubicin	1.00
Estramustine	0.93
Estreptonigrina	0.94
Etofoside	1.00
Etoglucid	0.99
Etoprine	0.38
Euparotin	1.00
FD-1	0.97
Fenafan	0.96
Fentirin	0.81
Fluarasquin	0.99
Fludarabine	0.95
Fludarabine phosphate	0.94
Fluorodofan	0.12
Fluorbenzotef	0.89
Fluorcitabine	0.01
Fluoromezin	0.46
Forfenimex	0.00
Formycin	0.96

(*continued on next page*)

Table 5 (*continued*)

Virtual screening of known anticancer compounds ($N = 213$, 87.7%)	
Name	*p*
Fotretamine	1.00
GANU	0.96
Gliocladic acid	1.00
Glutacyt	0.31
GP 48989	0.88
Hainanesine	0.99
Hainanolide	0.98
Hesferidin	1.00
Hexadepa	0.98
Hexafosfamid	0.96
Hexestrol di PO_4	0.96
Hisphen	0.87
Holocanthone	0.94
IDA	0.53
Idarubicin	1.00
IMET	0.88
IMET 3995	0.87
Indicine *N*-oxide	0.99
Iremycin	1.00
Irisquinone A	0.99
Irkutin	0.25
Juncusol	0.61
KW 2083	0.97
Lavendamycin	0.68
Leatrile	0.96
Leucodelphinidin	0.28
Leukogen	0.96
Lofenal	0.94
Loglutam-2	0.80
Lomenin-2	0.86
Lonin 4	1.00
Lumostine	0.99
Lymphamin	0.58
Lysepsin a	1.00
Lysepsin b	1.00
Mannomustine	0.99
Mannosulfan	0.99
Medorrubicin	1.00
Megoestrolacetat	1.00
Merophan	0.69
Mesyldegranol	1.00
Metamelfalam	0.57
Methopterine	1.00
Methotrexate	0.99
Metphol-B	0.95
Mitindomide	0.76
Mitobronitol	0.04
Mitomycin	0.99
Mitonafide	0.99
Mitopodozide	0.73
Mitotane	0.48
Mitoxantrone	0.98
Mycophenolicum acid	0.99
Nacaine	0.95
Neothramycin	0.37
Neplanocin C	0.95
Nicosin	0.96
Nifuron	0.68
Nimustine	0.98
Nitidine	0.87
Nitrazine	0.78
Nocodazole	0.47
NSC-143019	0.09
Octostanulum	1.00

Table 5 (*continued*)

Virtual screening of known anticancer compounds ($N = 213$, 87.7%)	
Name	*p*
Oxymatrine	0.99
Peucedanin	0.90
Pharsazin	0.96
Phenaline	1.00
Phenamet	1.00
Phenazinium	0.67
Phenester	0.70
Pidorubicin	1.00
Pifobroman	0.99
Piperazinedione	1.00
Piposulfan	0.99
Pretazettine	1.00
Pirazofurin	0.95
Porfirmycin	0.81
Procarbazine	0.94
Promicil	0.94
Psicoforanine	0.98
Pseudourea	0.80
Pterofterin	1.00
Plumbagin	0.14
Quinaspar	0.99
Rabdophillin G	1.00
Reumitsin	0.14
Ribofrine	1.00
RPCNU	0.98
Rutin-N-mustard	0.85
Semustine	0.99
Sesbanin	0.97
Solafalmitin	1.00
Sparsomycin	0.97
Spergualin	1.00
Spirogermanium	1.00
Stilbostat	0.97
Teroxirone	0.78
Thioguanosine	0.97
Thioinosine	0.96
Tiamiprine	0.27
Tiazicum acid	0.21
Tiazofurine	0.96
Tolnidamine	0.51
Toromycin	1.00
Trestolone acetate	1.00
Tretamine	0.87
Trichodermin	1.00
Triciribine	0.99
Trimetrexate	1.00
Trofosfamide	0.99
Tylophorine	1.00
Uramycin	0.97
V-100	0.93
Vinervine	0.80
Withaferin A	1.00

pounds are reported as active in the cited literature as active based on the results of many different tests. That reason compelled us to use LDA, instead of regression technique, to predict specific IC_{50} or any other quantitative value. So, the probabilities of activity predicted by our model have to be considered as the probability of a compound to be active in at least one of the many tests used in the original database. Consequencely, we have to probe the biological activity of the new compounds

Table 6. Results of cytotoxicity for assayed compounds

Compound	Predicted			Experimental		
	Probability	Predicted	Observed	IC_{50}[a] (μM)		
	%	Class	Class	L1210/0	CEM/0	Molt4/C8
4a	70.5	+	+	131.4 ± 53.3	149.2 ± 7.1	159.9 ± 53.3
4b	20.3	−	−	>400	>400	>400
4c	30.1	−	−	>400	>400	>400
5a	80.2	+	+	27.0 ± 4.8	27.2 ± 5.6	29.4 ± 4.2
5b	70.3	+	+	242.4 ± 22.7	185.6 ± 7.5	162.9 ± 11.3
5c	50.3	±	−	>400	>400	>400
Ara-A	91.5	+	+	14.3 ± 6.4	24.9 ± 1.9	11.9 ± 7.3

[a] 50% inhibitory concentration or compound concentration required to reduced proliferation of tumor cells by 50%.

at least in one test and that it is what we have done. In general, a very good coincidence between the theoretical prediction and the observed activity both for active and non-active compounds was observed. Compounds **4a**, **5a**, and **5b** became active, whereas **4b** and **4c** were shown to be non-active coinciding with model predictions. The compound **5c** was found to be non-active experimentally and presented a probability intermediate. The more interesting cytotoxic activity was found for compound **5a** with IC_{50} values between 27.0 and 29.4 μM for the three different cell lines assayed, which are comparable to those of Ara-A used as control, which was correctly predicted by the model with the highest probability.

The synthesis of all tested compounds was carried out by following the strategy given in Scheme 1. Compounds **4a**–**c** and **5a**–**c** were synthesized starting from the corresponding amino alcohol **1**, which was prepared in 20% overall yield from cyclohexene. The amino alcohol **1** was condensed with 2-amino-4,6-dichloropyrimidine in refluxing *n*-butanol containing triethylamine, providing the corresponding compound **2** in 60% yield. This compound was then coupled at position 5 with *p*-cholorobenzenediazonium chloride giving the diazo compound, which, after cleavage of the diazo linkage with Zn in an acidic medium, gave **3** in 30% yield from **2**. Purinyl analogue **4a** was obtained in 60% yield by treating **3** with triethyl orthoformate in hydrochloric acid. Diazotization of **3** with sodium nitrite in acetic acid and spontaneous cyclization of the intermediate diazonium salt gave 8-azapurinyl analogue **5a** in 80% yield. Nucleophilic substitution of the 6-chloro substituents of **4a** and **5a** was achieved by treatment with sodium hydroxide giving guanine derivative **4b** and 8-azaguanine derivative **5b** in 71% and 92% yields, respectively. Similarly, reaction of **4a** and **5a** with liquid ammonia gave, respectively, 2,6-diaminopurinyl derivative **4c** in 40% yield and 8-aza-2,6-diaminopurinyl derivative **5c** in the same yield.

2.4. The back-projection of the QSAR

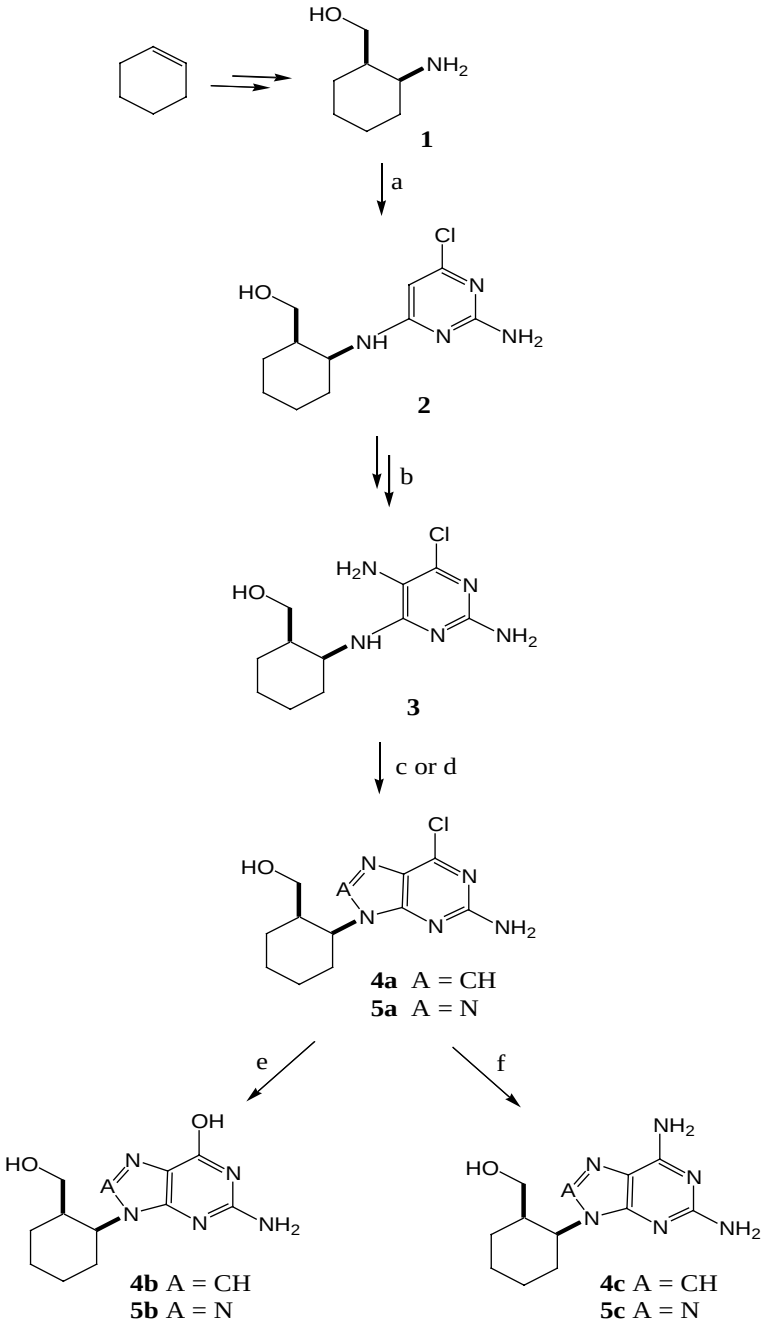
Many 3D techniques generate an output that is easily visualized and can be used as an "idea generator" for new drug candidates. This visualization can be described as a back-projection of the model into the original molecular space. In this sense, the molecular descriptors can be split into two groups: not back-projectable ones and back-projectable 3D molecular descriptors.[51] Recently, our research group has been working for the first time on the back-projection of 2D and chiral topological indices, which are derived from chemical graphs weighted with 3D information.[28,52–54] In this work, we carried out the back-projection analysis by directly substituting the numerical values calculated for the entropies for every group of atoms in the QSAR Eq. 1. Figure 2 depicts the results of back-projection analysis for the more active compound reported here. At first view, it highlights the higher contribution of the purine ring (45.5), of chlorine atoms (10.3), and of hydroxymethyl group (13.2) coinciding with experimental findings.[55] This fact, is related to the positive contribution of the *cis*-substituted C1–C2 bond of the carbocycle (11.7).[56] Finally, the low negative contribution of the other carbon atoms (−3.2) directly points to the aliphatic ring as a chemical feature susceptible to synthetic modification in future works.

3. Experimental

3.1. Chemistry

Melting points were determined with a Stuart Scientific melting point apparatus and are uncorrected. IR spectra were recorded on a Perkin-Elmer 1640 FT spectrophotometer (KBr disks, ν in cm^{-1}). ^{1}H and ^{13}C NMR spectra were recorded on Bruker DPX (250 MHz) and Bruker AMX (500 MHz) spectrometers, using TMS as internal standard (chemical shifts as δ in ppm, J in Hz). Mass spectra and HRMS (EI) were obtained using a Hewlett Packard 5988A spectrometer and Micromass Autospec spectrometer, respectively. Silica gel (Merck 60, 230–400 mesh) was used for flash chromatography (FC). Analytical thin layer chromatography (TLC) was performed on plates precoated with silica gel (Merck 60 F254, 0.25 mm).

3.1.1. (±)-*cis*-9-[2-(Hydroxymethyl)cyclohexyl]guanine (4b). A mixture of **4a** (60 mg; 0.21 mmol) and 0.33 M NaOH (3 mL) was refluxed for 5 h. The solvent was then evaporated under vacuum (azeotropic mixture with EtOH–toluene) and the solid residue was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **4b** (40 mg; 71%): mp 280–282 °C.



Scheme 1. Reagents and conditions: (a) 2-amino-4,6-dichloropyrimidine, Et_3N, *n*-BuOH, reflux 24 h, 60%; (b) *p*-chloroaniline, $NaNO_2$, 12 M HCl 0 °C; then Zn, AcOH, EtOH, reflux 1 h, 30%; (c) 12 M HCl, $CH(OEt)_3$, reflux 12 h, 60% (**4a**); (d) $NaNO_2$, H_2O, AcOH, 80% (**5a**); (e) 0.33 M NaOH, reflux 5 h, 71% (**4b**), 92% (**5b**); (f) NH_4OH, reflux, 40% (**4c** and **5c**).

^{1}H NMR (DMSO-d_6): 10.5 (br s, 1H, OH aromatic), 7.71 (s, 1H, H-8), 6.46 (br s, 2H, NH_2), 4.36 (m, 2H, CH-N + OH aliphatic), 3.45 (m, 1H, *H*CH-O), 3.00 (m, 1H, HC*H*-O), 2.20–1.30 (m, 9H, $(CH_2)_4$ + CH-C-O). ^{13}C NMR (DMSO-d_6): 157.2 (6), 153.8 (4), 152.3 (2), 136.3 (8), 116.9 (5), 58.3 (7′), 54.0 (1′), 40.3 (6′), 27.0, 25.7, 24.9, 20.7. IR: 3337, 3123, 2931, 1685, 1653, 1370. MS *m*/*z* (%): 263 (M^+, 16), 233 (M^+-CH_2O, 6), 152 (M^+-$C_7H_{11}O$, 38), 151 (M^+-$C_7H_{12}O$, 100), 109 (35), 67 (11). HRMS calcd for $C_{12}H_{17}N_5O_2$: 263.1382, found: 263.1374.

3.1.2. (±)-*cis*-2,6-Diamino-9-[2-(hydroxymethyl)cyclohexyl]-purine (4c). A solution of **4a** (60 mg; 0.21 mmol) in NH_4OH (8.3 mL) was refluxed for 6 h. The solvent was evaporated under vacuum and the residue was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **4c** (22 mg; 40%): mp 217–219 °C.

^{1}H NMR (DMSO-d_6): 7.74 (s, 1H, H-8), 6.68 (br s, 2H, NH_2), 5.81 (br s, 2H, NH_2), 4.45 (m, 2H, CH-N + OH aliphatic), 3.35 (m, 1H, *H*CH-O), 2.97 (m, 1H, HC*H*-O), 2.17–1.23 (m, 9H, $(CH_2)_4$ + CH-C-O). ^{13}C NMR

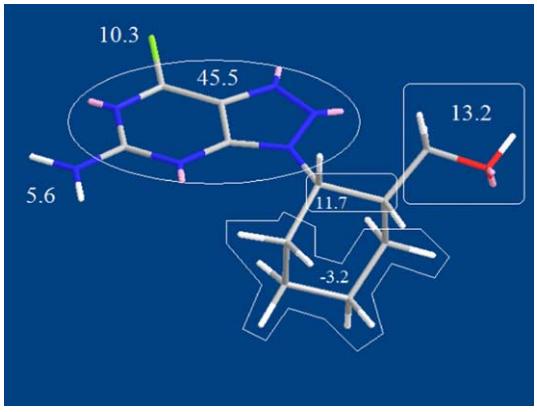


Figure 2. Back-projection analysis for the more active compound found.

(DMSO-d_6): 160.2 (6), 156.3 (4), 152.1 (2), 136.3 (8), 113.4 (5), 58.6 (7′), 53.5 (1′), 40.4 (6′), 27.3, 25.5, 24.9, 20.9. IR: 3297, 3156, 2920, 2856, 1664, 1595, 1477, 1450, 1397. MS *m*/*z* (%): 262 (M^+, 14), 232 (M^+-CH_2O, 8), 177 (18), 151 (M^+-$C_7H_{11}O$, 23), 150 (M^+-$C_7H_{12}O$, 100), 108 (19), 67 (3). HRMS calcd for $C_{12}H_{18}N_6O$: 262.1542, found: 262.1531.

3.1.3. (±)-*cis*-2-Amine-8-aza-6-chloro-9-[2-(hydroxymethyl)cyclohexyl]-9H-purine (5a). $NaNO_2$ (42 mg; 0.61 mmol) was added to a solution of **3** (130 mg; 0.48 mmol) in AcOH (0.75 mL) and H_2O (1.1 mL) at 0 °C. The mixture was stirred at 0 °C for 1 h. The solvent was evaporated under vacuum and the residue was purified by FC with 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **5a** (108 mg, 80%): mp 203–205 °C.

^{1}H NMR (DMSO-d_6): 7.62 (br s, 2H, NH_2), 4.95 (m, 1H, CH-N), 4.29 (m, 1H, OH aliphatic), 3.27 (m, 1H, *H*CH-O), 2.98 (m, 1H, HC*H*-O), 1.54-1.45 (m, 9H, $(CH_2)_4$ + CH-C-O). ^{13}C NMR (DMSO-d_6): 162.0 (6), 152.1 (4), 151.9 (2), 127.9 (5), 60.6 (7′), 55.1 (1′), 41.7 (6′), 28.9, 24.9, 23.1, 22.6. IR: 3367, 3198, 2910, 1646, 1607, 1568, 1508, 1004. MS *m*/*z* (%): 284 ($[M+2]^+$, 4), 282 (M^+, 15), 254 ($[M+2]^+$-CH_2O, 3), 252 (M^+-CH_2O, 9), 224 (9), 173 ($[M+2]^+$-$C_7H_{12}O$, 14), 171 (M^+-$C_7H_{11}O$, 100), 170 (M^+-$C_7H_{12}O$, 24), 144 (21), 79 (21), 67 (10). HRMS calcd for $C_{11}H_{15}ClN_6O$: 282.0996, found: 282.0994.

3.1.4. (±)-*cis*-8-Aza-9-[(2-hydroxymethyl)cyclohexyl]guanine (5b). Prepared from **5a** (70 mg; 0.25 mmol) in the same way as for **4b** from **4a**.

5b (60 mg; 92%): mp 204–206 °C.

^{1}H NMR (DMSO-d_6): 11.01 (s, 1H, OH aromatic), 7.03 (br s, 2H, NH_2), 5.24 (br s, 1H, OH aliphatic), 4.78 (m, 1H, CH-N), 3.20 (m, 1H, *H*CH-O), 2.95 (m, 1H, HC*H*-O), 2.06–1.15 (m, 9H, $(CH_2)_4$ + CH-C-O). ^{13}C NMR (DMSO-d_6): 156.5 (6), 155.6 (4), 151.8 (2), 124.5 (5), 60.5 (7′), 54.3 (1′), 41.8 (6′), 29.0, 24.8, 23.1, 22.7. IR: 3292, 2925, 1706, 1659, 1579, 1387. MS *m*/*z* (%): 264 (M^+,15), 234 (M^+-CH_2O, 7), 219 (5), 179 (8), 153 (M^+-$C_7H_{11}O$, 100), 152 (M^+-$C_7H_{12}O$, 35), 126 (40), 67 (11), 55 (13). HRMS calcd for $C_{11}H_{16}N_6O_2$: 264.1335, found: 264.1334.

3.1.5. (±)-*cis*-2,6-Diamino-8-aza-9-[2-(hydoxymethyl)cyclohexyl]purine (5c). $NaNO_2$ (41 mg; 0.60 mmol) was added to a solution of **3** (130 mg; 0.48 mmol) in 1 M HCl (1 mL) at 0 °C and the mixture was stirred for 15 min. Then NH_4OH (2.2 mL) was added and the mixture was refluxed for 5 min. After cooling, the precipitate was recovered by filtration and purified by FC using 95:5 CH_2Cl_2/CH_3OH as eluent, which gave pure **5c** (50 mg; 40%): mp 252–254 °C.

^{1}H NMR (DMSO-d_6): 7.50 (s, 2H, NH_2), 6.37 (s, 2H, NH_2), 4.80 (m, 1H, CH-N), 4.43 (m, 1H, OH aliphatic), 3.22 (m, 1H, *H*CH-O), 2.90 (m, 1H, HC*H*-O), 2.07–1.40 (m, 9H, $(CH_2)_4$ + CH-C-O). ^{13}C NMR (DMSO-d_6): 163.1 (6), 156.7 (4), 151.6 (2), 119.8 (5), 60.6 (7′), 54.1 (1′), 42.1 (6′), 29.0, 24.7, 23.2, 22.9. IR: 3330, 3184, 2933, 1670, 1653, 1636, 1604, 1479, 1419. MS (*m*/*z*): 263 (M^+, 17), 246 (M^+-NH_3, 9), 233 (M^+-CH_2O, 14), 232 (M^+-CH_2OH, 22), 181 (32), 164 (21), 152 (M^+-$C_7H_{11}O$, 100), 125 (59), 67 (19), 53 (11). HRMS calcd for $C_{11}H_{17}N_7O$: 263.1495, found: 263.1486.

3.2. Computational approach

The calculation of Θ_k has been largely explained in the literature, consequently only a brief definition is given as follows (please see the references for details).[24–27,38,57] The entropy $\Theta_k(j)$ is calculated summing up atom by atom logarithmic terms dependent on the absolute probability $^A\pi_k(j)$ with which every atom j attracts the electrons of its neighbors placed at distance k in the molecule

$$\Theta_k = -\sum_{j=1}^{n} \Theta_k(j) = -k_B \sum_{j=1}^{n} {}^A\pi_k(j) \cdot \ln {}^A\pi_k(j). \qquad (2)$$

All computations were carried out on a PC Pentium 4 2.6 GHz. The calculation of Θ_k for any organic molecule was implemented in the software MARCH-INSIDE[58] developed in our laboratory. This software has a graphical interface that makes it user-friendly for medicinal chemists. In Eq. 1, the coefficients of the classification function were determined by the forward stepwise LDA[42] method implemented on STATISTICA 6.0.[43] No variable reduction method was applied; it was not necessary in this particular case, we calculated only 6 molecular descriptors, $k = 0, 1, \ldots, 5$. Individual compounds were assigned to the training and test sets at random, with random numbers taken from a random number table.

3.3. Anti-cancer 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%.[59]

References and notes

1. Green, S.; Goldberg, A.; Zurlo, J. *Toxicol. Sci.* **2001**, *63*, 6.
2. *Progress in the Reduction and Refinement and Replacement of Animal Experimentation*; Balls, M., Van Zeller, A.-M., Halder, M. E., Eds.; Elselvier: Amsterdam, 2000.
3. Denny, W. A. The Role of Medicinal Chemistry in the Discovery of DNA-Active Anticancer Drugs: From Random Searching, Through Lead Development, to the Novo Design. In *The Search for Anticancer Drugs*; Waring, M. J., Ponder, B. A. J., Eds.; Kluwer Academy Publishers: Dordrecht, 1992; Chapter 2.
4. Pevarello, P.; Brasca, M. G.; Amici, R.; Orsini, P.; Traquandi, G.; Corti, L.; Piutti, C.; Sansonna, P.; Villa, M.; Pierce, B. S.; Pulici, M.; Giordano, P.; Martina, K.; Fritzen, E. L.; Nugent, R. A.; Casale, E.; Cameron, A.; Ciomei, M.; Roletto, F.; Isacchi, A.; Fogliatto, G. P.; Pesenti, E.; Pastori, W.; Marsiglio, A.; Leach, L. K.; Clare, P. M.; Fiorentini, F.; Varasi, M.; Vulpetti, A.; Warpehoski, M. A. *J. Med. Chem.* **2004**, *47*, 3367.
5. Saczewski, F.; Reszka, R.; Gdaniec, M.; Grünert, R.; Bednarski, P. J. *J. Med. Chem.* **2004**, *47*, 3438–3449.
6. Glenn, M. P.; Kahnberg, P.; Boyle, G. M.; Hansford, K. A.; Hans, D.; Martyn, A. C.; Parsons, P. G.; Fairlie, D. P. *J. Med. Chem.* **2004**, *47*, 2984.
7. Lunney, E. A. *Med. Chem. Res.* **1998**, *8*, 352.
8. Zauhar, R. J.; Moyna, G.; Tian, F. L.; Li, Z. J.; Welsh, W. J. *J. Med. Chem.* **2003**, *46*, 5674.
9. Pan, D.; Liu, J.; Senese, C.; Hopfinger, A. J.; Tseng, Y. *J. Med. Chem.* **2004**, *47*, 3075.
10. Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47*, 2499.
11. Melani, F.; Gratteri, P.; Adamo, M.; Bonaccinin, C. *J. Med. Chem.* **2003**, *46*, 1359.
12. Kier, L. B.; Hall, L. H. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach Sci.: Amsterdam, 1999; pp 455–489.
13. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
14. Lepper, E. R.; Ng, S. S. W.; Gütschow, M.; Weiss, M.; Hauschildt, S.; Hecker, T. K.; Luzzio, F. A.; Eger, K.; Figg, W. D. *J. Med. Chem.* **2004**, *47*, 2219.
15. Yao, S.-W.; Lopes, V. H. C.; Fernandez, F.; Garcia-Mera, X.; Morales, M.; Rodriguez-Borges, J. E.; Cordeiro, M. N. D. S. *Bioorg. Med. Chem.* **2003**, *11*, 4999.
16. Babu, M. S.; Denny, W. A.; Kurup, A.; Hansch, C. *Bioorg. Med. Chem.* **2001**, *9*, 2757.
17. Garg, R.; Denny, W. A.; Hansch, C. *Bioorg. Med. Chem.* **2000**, *8*, 1835.
18. Xie, A.; Liao, Ch.; Li, Z.; Ning, Z.; Hu, W.; Lu, X.; Shi, L.; Zhou, J. *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 273.
19. David-Cordonnier, M.-H.; Laine, W.; Gaslonde, T.; Michel, S.; Tillequin, F.; Koch, M.; Leonce, S.; Pierre, A.; Bailly, C. *Curr. Med. Chem. Anti-Cancer Agents* **2004**, *4*, 83.
20. Venkatachalam, T. K.; Goodman, P. A.; Qazi, S.; D'Cruz, O.; Uckun, F. M. *Curr. Pharm. Des.* **2004**, *10*, 1713.
21. J.A. Freund, T. Poschel, Eds.; *Stochastic Processes in Physics, Chemistry, and Biology*. In: Lect. Notes Phys. Springer-Verlag, Berlin, Germany, 2000.
22. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernández, S. I.; Morales, A.; Serrano, H. S.; González, J.; Ramos de Armas, R. *J. Mol. Mod.* **2002**, *8*, 237.
23. González-Díaz, H.; Ramos de Armas, R.; Molina, R. R. *Bull. Math. Biol.* **2003**, *65*, 991.
24. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, E.; Nasco, O.; Uriarte, E.; Castañedo, C. N.; Cabrera-Pérez, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
25. González-Díaz, H.; Ramos de Armas, R.; Molina, R. *Bioinformatics* **2003**, *19*, 2079.
26. Ramos de Armas, R.; González-Díaz, H.; Molina, R.; Uriarte, E. *Prot. Struct. Funct. Bioinf.* **2004**, *56*, 715.
27. González-Díaz, H.; Molina, R.; Uriarte, E. *Polymers* **2004**, *45*, 3845.
28. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos de Armas, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Mod.* **2003**, *9*, 395.
29. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de Armas, R. *Bull. Math. Biol.* **2004**, *66*, 1285.
30. González-Díaz, H.; Hernández, S. I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
31. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1573.
32. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 2881.
33. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sánchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 1005.
34. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, M.; Mayón-Peréz, M.; García-Sánchez, R. N. *Bioorg. Med. Chem.* **2005**, *13*, 1293.
35. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg. Med. Chem.* **2005**, *13*, 1119–1129.
36. González-Díaz, H.; Agüero, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castañedo, N. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 551.
37. González-Díaz, H.; Uriarte, E.; Ramos de Armas, R. *Bioorg. Med. Chem.* **2005**, *13*, 323.
38. González-Díaz, H.; Tenorio, E.; Castañedo, N.; Santana, L.; Uriarte, E. *Bioorg. Med. Chem.* **2005**, *13*, 1523.
39. Shannon, C. E. *The Mathematical Theory of Communication*; University of Illinois Press: Urbana, 1955.
40. Negwer, M. *Organic Chemical Drugs and Their Synonyms*; Akademie-Verlag: Berlin, 1987.
41. Kleeman, A.; Engel, J.; Kutscher, B.; Reichert, D. *Pharmaceutical Substances*, 4th ed.; George Thieme Verlag: Stuttgart, 2001.
42. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, R., Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, pp 265–282.
43. Statsoft. Inc. STATISTICA. 2002, Version. 6.0.
44. Randić, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
45. Randić, M. *N. J. Chem.* **1991**, *15*, 517.
46. Randić, M. *J. Mol. Struct. (TEOCHEM)* **1991**, *233*, 45.
47. Randić, M. *J. Comput. Chem.* **1993**, *4*, 363.
48. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
49. Gálvez, J.; García-Domenech, R.; Gómez-Lechón, M. J.; Castell, J. V. *J. Mol. Struct. (TEOCHEM)* **2000**, *504*, 241.

50. Benigni, R.; Cotta-Ramusino, M.; Gallo, G.; Giorgi, F.; Giuliani, A.; Vari, M. R. *J. Med. Chem.* **2000**, *43*, 63699.
51. Stief, N.; Baumann, K. *J. Med. Chem.* **2003**, *46*, 1390.
52. Cabrera-Pérez, M. A.; González-Díaz, H.; Fernández-Teruel, C.; Plá-Delfina, J. M.; Bermejo-Sanz, M. *Eur. J. Phram. Biopharm.* **2002**, *53*, 317.
53. Estrada, E.; González-Díaz, H. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 75.
54. Marrero-Ponce, Y. *Molecules* **2003**, *8*, 687.
55. Hattori, H.; Tanaka, M.; Fukushima, M.; Sasaki, T.; Matsuda, A. *J. Med. Chem.* **1996**, *39*, 5005.
56. Santana, L.; Teijeira, M.; Uriarte, E.; Balzarini, J.; De Clercq, E. *Eur. J. Med. Chem.* **2002**, *37*, 755.
57. González-Díaz, H.; Zaís-Urra, L.; Molina, R.; Uriarte, E. *Polymer* **2005**, *46*, 2791.
58. González-Díaz, H.; Molina, R.; Hernández, I. MARCH-INSIDE® 2002, version 2.0. Information request to: gonzalezdiazh@yahoo.es.
59. 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.