

# Design, Synthesis, and Evaluation of Antineoplastic Activity of Novel Carbocyclic Nucleosides

Aliuska M. Helguera,<sup>[a, c, d]</sup> J. E. Rodríguez-Borges,<sup>[b]</sup> Olga Caamaño,<sup>[e]</sup> Xerardo García-Mera,<sup>\*,[e]</sup> Maykel Pérez González,<sup>[c]</sup> and M. Natália D. S. Cordeiro<sup>\*,[a]</sup>

This paper is dedicated to the memory of our dear colleague, Maykel Pérez González, who unfortunately has passed away very suddenly last year.

**Abstract:** Cancer is the leading cause of death among men and women under age 85. Every year, millions of individuals are diagnosed with cancer. But finding new drugs is a complex, expensive, and very time-consuming task. Over the past decade, the cancer research community has begun to address the in silico modeling approaches, such as Quantitative Structure-Activity Relationships (QSAR), as an important alternative tool for targeting potential anticancer drugs. With the compilation of a large dataset of nucleosides synthesized in our laboratories, or elsewhere, and tested in a single cytotoxic assay under the same experimental conditions, we recognized a unique opportunity to attempt to build predictive QSAR models. Early efforts with 2D classification models built from part of this dataset were very encouraging. Here we report a further detailed evaluation of classification models to flag potential anticancer activities derived from a variety of 3D molecular representa-

tions. A quantitative 3D-model model that discriminates anticancer compounds from the inactive ones was attained, which allowed the correct classification of 82% of compounds in such a large and diverse dataset, with only 5% of false inactives and 11% of false actives. The model developed here was then used to select and design a new series of nucleosides, by classifying beforehand them as active/inactive anticancer compounds. From the compounds so designed, 22 were synthesized and evaluated for their inhibitory effects on the proliferation of murine leukemia cells (L1210/0), of which 86% were well-classified as active or inactive, and only two were false actives, corroborating the good predictive ability of the present discriminant model. The results of this study thus provide a valuable tool for the design of novel potent anticancer nucleoside analogues.

**Keywords:** Antitumor agents · Nucleosides · QSAR · 3D-DRAGON descriptors · Leukemia

## 1 Introduction

Cancer is one of the major causes of disease and death all over the world.<sup>[1]</sup> For instance, in Europe, an estimated 3.2 million people were diagnosed with cancer in 2006, there were 1.7 million cancer deaths, and the annual number of new cases has increased dramatically.<sup>[1]</sup> Oncology has thus become a leading therapeutic area for pharmaceutical research with ca. 650 drugs undergoing clinical trials (www.pharma.org). However, finding new drugs is a complex, expensive, and very time-consuming task. Indeed, bringing a new drug to the market can take up to 15 years and cost around €620 million.<sup>[1]</sup> Lately, rational drug-design strategies, such as Quantitative Structure-Activity Relationships (QSAR) modeling methods, have emerged as a promising alternative or complementary tool towards the effective screening of potential drugs, being thus increasingly attracting the attention of medicinal chemists and of the pharmaceutical industry. QSAR modeling may be better regarded as an exercise to filter drug candidates, before they are subjected to more intensive calculations such as dock-

[a] A. M. Helguera, M. N. D. S. Cordeiro  
REQUIMTE, Chemistry Department, University of Porto  
Rua do Campo Alegre 687, 4169-007 Porto, Portugal  
fax: +351 220402659  
\*e-mail: ncordeir@fc.up.pt

[b] J. E. Rodríguez-Borges  
CIQ, Chemistry Department, University of Porto  
Rua do Campo Alegre 687, 4169-007 Porto, Portugal

[c] A. M. Helguera, M. P. González  
CBQ, Department of Chemistry, Central University of Las Villas  
Santa Clara, 54830, Villa Clara, Cuba

[d] A. M. Helguera  
Department of Chemistry, Central University of Las Villas  
Santa Clara, 54830, Villa Clara, Cuba

[e] O. Caamaño, X. García-Mera  
Department of Organic Chemistry, Faculty of Pharmacy,  
University of Santiago de Compostela  
15706 Santiago de Compostela, Spain  
fax: +34 981594912  
\*e-mail: xerardo.garcia@usc.es

 Supporting Information for this article is available on the WWW under <http://dx.doi.org/10.1002/minf.200900033>

ing or an experimental measurement of activity (in vitro) and under real conditions (in vivo) lastly.

The nucleoside analogues (NAs) were among the first chemotherapeutic agents introduced for the medical treatment of cancer.<sup>[2]</sup> One of kind of NAs that have been synthesized for fighting cancer are carbocyclic nucleosides,<sup>[3]</sup> in which the methylene group replaces the oxygen atom of a furanose ring. As such, the glycosidic bond is resistant to nucleoside phosphorylases and hydrolases, making the carbocyclic nucleosides more stable towards metabolic degradation.<sup>[4]</sup> Interestingly, in some cases, the substitution of the sugar ring by a carbocyclic ring does not affect the enzyme recognition (especially kinases as target enzymes).<sup>[5]</sup> Due to these features, carbocyclic nucleosides have received much attention as potential chemotherapeutic agents,<sup>[2]</sup> such as abacavir and entecavir (Figure 1).

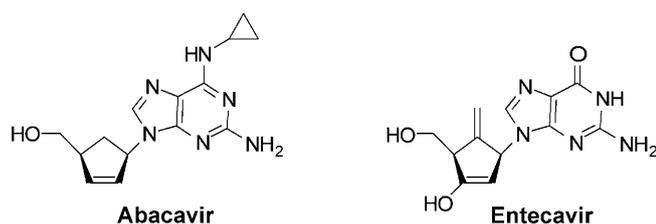


Figure 1. Important carbocycle nucleosides.

For years our research group has been engaged in the design, synthesis and evaluation of nucleoside analogues in an effort to develop potential anticancer or antiviral agents<sup>[4,5]</sup>. We recognized an excellent opportunity when we could assemble a set of over 200 NA compounds, previously synthesized in one of our laboratories<sup>[3,6-9]</sup> or elsewhere<sup>[10-31]</sup>, all measured in a single, consistent cytotoxic assay against murine leukemia L1210/0 cells<sup>[32]</sup>. With the desire to build statistical sound and predictive models from such data, we have recently presented a detailed QSAR analysis covering the most important two-dimensional (2D) structural features that rule the anticancer activity within NAs.<sup>[33]</sup> The 2D-structural information then gathered and the QSAR model per se can well aid to discriminate active/inactive NAs. That was confirmed by the classification results obtained with an external test set comprising 20 diverse carbonucleoside analogues. Indeed, for that external set, the percentage of overall discrimination attained was quite impressive (75%), given the diversity of the test set and the complexity of the biologic response being modeled. Finally, a clustering search analysis, to identify the similarities to natural nucleosides of the well-predicted active NAs, and a structural interpretation of the results, taking into account the mechanism of action responsible for their cytotoxic activity, has been carried out. The information provided by this analysis showed us that, in particular, the indan derivatives are closely related to adenosine and guanosine.<sup>[33]</sup>

The 2D representation of molecules can encode important information on adjacency, branching and relative distance among different functionalities in a numerical form that determines a wide range of physicochemical and biological properties, but it does not take into account information concerning conformational aspects, i.e. bond lengths, bond angles and torsion angles.<sup>[34]</sup> Besides, recognition of the importance of the three-dimensional (3D) structure and stereochemistry of molecules to their biological activity, increasing knowledge of the 3D structure of biological macromolecules such as proteins, and awareness of the limitations of *classical* 2D approaches, led to many attempts to generate 3D descriptors either for complementing 2D-QSAR models or for standalone 3D-QSAR models. In this work, we systematically examined the use of 3D descriptors along with linear discriminant analysis for probing the anticancer activity of NAs. This strategy was found to produce a final discriminant QSAR model that exhibits very good cross-validation statistics and additionally, perform well on an external test set comprising other NA chemicals designed by us and with unknown activity. The design, synthesis, and biological evaluation of this new series of carbonucleosides are reported.

## 2 Materials and Methods

### 2.1 QSAR Modeling

#### 2.1.1 Data Set

All the compounds used here are primarily nucleoside analogues, derived from purinic and pyrimidinic bases, and were experimentally assayed for their inhibitory effects ( $IC_{50}$ ) in the proliferation of L1210/0 cancer cells. These experiments have been conducted at the Rega Institute for Medical Research of the Katholieke Universiteit Leuven in Leuven, Belgium, following the same in vitro assay protocol.<sup>[26]</sup> One can rely on the *quality* of such biological data, which has been measured by a single protocol, at the same laboratory, by even the very same staff.

The compounds were firstly clustered into two groups according to their  $IC_{50}$  values. The first group – *actives* – includes all chemicals with  $IC_{50} < 200 \mu\text{M}$ , while the second one – *inactives* – includes those with  $IC_{50} \geq 200 \mu\text{M}$ . This classification criterion was adopted not only because over that concentration chemicals can be too toxic and therefore lack biological value but also to get a reasonable ratio of active/inactive chemicals in the dataset. We have also discarded all chemicals with disconnected structures like salts and polymers. Finally we managed to assemble a large, balanced dataset of 300 chemicals comprising 107 actives and 193 inactives.

A necessary but delicate task in any QSAR modeling is predictive validation, i.e. to assess model adequacy for new compounds. In general, the most reliably way to predictively validate a model is by external validation, which consists

of making predictions for an independent set of data not used in the model setup. Here we select a small subset (22 compounds) of the NAs after a design process to act as an *external test* set. These compounds have been synthesized in one of our laboratories following the procedures described above or reported previously by us,<sup>[35–37]</sup> and four of them have already been assayed experimentally in L1210 cells and their activity reported,<sup>[36]</sup> while the others were evaluated here for the first time. A complete list of the training set (278 compounds) along with the reported experimental cytotoxicity ( $IC_{50}$  values expressed in  $\mu\text{M}$ ) is given as Supporting Information.

### 2.1.2 Molecular Descriptors

Our models are based on six different blocks of 3D-descriptors that are available in the DRAGON software package<sup>[38a]</sup>, with a successful history in structure–activity and structure–property correlation. They include:

(i) *Randić molecular profiles*

A set of 41 descriptors derived from the distance distribution moments of the geometry matrix, including both molecular as shape profiles.

(ii) *Geometrical descriptors*

A set of 74 descriptors consisting of different kinds of conformation-dependent descriptors based on molecular geometry.

(iii) *Radial Distribution Function (RDF) descriptors*

This set consists of 135 descriptors calculated from the radial distribution function, that is to say, the probability distribution to find an atom in a spherical volume centered in each atom.

(iv) *Molecule Representation of Structures based on Electron diffraction (3D-MoRSE) descriptors*

The 3D-MoRSE approach considers the molecular information derived from an equation used in electron diffraction studies. The 160 3D-MoRSE descriptors are calculated by summing atomic weights viewed by different angular scattering functions.

(v) *Weighted Holistic Invariant Molecular (WHIM) descriptors*

These indices are calculated by a principal component analysis on the centered Cartesian coordinates of the atoms by using a weighted covariance matrix. Two kinds of WHIM descriptors have been defined: 66 *directional* and 33 *nondirectional* descriptors.

(vi) *GEometry, Topology, and Atoms-Weighted Assembly (GETAWAY) descriptors*

The GETAWAY descriptors are based on a leverage matrix, similar to the one defined in statistics and employed for regression diagnostics, called molecular influence matrix (*MIM*). These descriptors are divided into two different sets: the first set consists of 104 *H-GETAWAY* descriptors, derived by using only the information provided by the *MIM*, and the remaining 93 *R-GETAWAY* descriptors,

derived by joining the *MIM* information with the geometric interatomic distances in the molecule.

3D-molecular descriptors were computed after fully optimizing the geometry of each molecule by the semi-empirical quantum-mechanics method, Austin Model 1 (AM1), implemented in MOPAC program<sup>[38b]</sup>. By disregarding descriptors with constant or near constant values inside each class, a final subset of 671 descriptors was then used for building the QSAR models.

### 2.1.3 Statistical Methods

Linear discriminant analysis (LDA), as implemented in the STATISTICA software (version 7.0),<sup>[38c]</sup> was applied here to find classification models (Eqs. 1–2) that best describe the cytotoxic activity  $P$ , as a linear combination of the predictor  $X$ -variables (3D descriptors).

In developing the models,  $P$  values of +1 and –1 were assigned to active and inactive compounds, respectively, but a posteriori probabilities are used instead to assert the models' classification of compounds. In particular, when the probability of being active did not differ more than 5% from that of being inactive, the case was considered as unclassified (U) by the model.

The forward stepwise (FS) technique was applied to select the molecular descriptors ( $X$ -variables) with the highest influence on the anticancer activity. This technique begins by including the variable which yields the best linear fit in terms of explaining the response. The next variable is included as that variable which most significantly improves the existing model. Once this new model is determined, the variables included are tested to see if the model can be improved by dropping them from the model. If the model can be improved, the variable is removed and the stepwise procedure is repeated until no further variables are either included or removed.

As the variables included in the first developed model (Eq.1) were strongly interrelated to each other, this may well lead to a multicollinearity problem, and can cause problems in interpreting the individual estimated coefficients. One very useful and informative approach of avoiding multicollinearity is the orthogonalization procedure suggested by Randić some year's ago.<sup>[39]</sup> In such a procedure, after choosing a starting descriptor, subsequent descriptors are added only as their orthogonal complements to the descriptors already present. This has the advantages that the equation coefficients are stable (i.e., they do not change as new descriptors are added), and the new information supplied by each additional descriptor is clearly distinguished in the final equation statistics. Here, to deal with the multicollinearity problem, we resorted to Randić's procedure and orthogonalized the variables following the order selected by the FS scheme (Mor12m, ITH, ADDD, MEcc, Mor04e, E3s, Mor19u, J3D, Au, and RDF120v). The resulting orthogonal-descriptor model was standardized afterwards.

## 3 Experimental

### 3.1 Chemistry

#### 3.1.1 General

Silica gel was purchased from Merck. All other chemicals used were of reagent grade and were obtained from Aldrich Chemical Co. The reactions were monitored by thin layer chromatography conducted on E. Merck TLC plates (silica gel 60 F-254, aluminum back) and visualized with UV light or iodine or anisaldehyde solution. The products were purified by column chromatography on silica gel (Merck 60, 230–240 mesh). Melting points were determined on a Reichert Kofler Thermopan or in capillary tubes on a Büchi 510 apparatus and are uncorrected. Infrared spectra were recorded on a Perkin-Elmer 1640-FT spectrophotometer and the main bands are given in  $\text{cm}^{-1}$ .  $^1\text{H}$  NMR (300 MHz) and  $^{13}\text{C}$  NMR spectra (75.47 MHz) were recorded on a Bruker WM AMX spectrometer using TMS (tetramethylsilane) as an internal standard (chemical shifts ( $\delta$ ) in parts per million, J in hertz). Mass spectra were performed on a Hewlett-Packard HP5988A mass spectrometer by electron impact (EI), or on a Finnigan Trace-MS mass spectrometer by chemical ionization (CI). Optical rotations at the sodium D-line were determined using a Perkin-Elmer 241 thermostated polarimeter. Elemental analyses were obtained on a Perkin-Elmer 240B microanalyser by the Microanalysis Service of the University of Santiago de Compostela. GLC analyses were carried out on a Hewlett-Packard 5890 II apparatus provided with a flame ionization detector, using a semicapillary column (5 m  $\times$  0.53 mm i.d., film thickness 2.65  $\mu\text{m}$ ) and helium as carrier gas. The purity of the compounds used on the biological tests was at least 95% and was determined by combustion analysis. (1*S*,2*R*)-(cis)-1-Amino-2-indanol, 99% was purchased on Aldrich. Inc.

**(+)-(1*R*,cis)-3-[(5-Amino-6-chloropyrimidin-4-ylaminomethyl)-2,2-dimethylcyclobutyl]methanol (2a).** A mixture of freshly prepared amino alcohol **1a**<sup>[42]</sup> (3.0 g, 20.98 mmol), and 5-amino-4,6-dichloropyrimidine (4.0 g, 24.4 mmol) in  $\text{Et}_3\text{N}$  (13 mL) and *n*-BuOH (83 mL) was refluxed under argon for 72 h, the solvents were removed under reduced pressure, and the resulting oil was chromatographed (silica gel  $\text{CH}_2\text{Cl}_2/\text{MeOH}$ , 25:1). Removal of the solvent left **2a** (5.11 g, 90%) as a white solid. Mp 181–183 °C ( $\text{Et}_2\text{O}/\text{EtOH}$ ).  $[\alpha]_{\text{D}}^{25} + 13.02$  (c 0.5, MeOH). IR (KBr):  $\nu = 3196, 2951, 1654, 1577, 1458, 1423, 1225, 920 \text{ cm}^{-1}$ .  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta = 0.96$  (3H, s,  $\text{CH}_3$ ), 1.04 (s, 3H,  $\text{CH}_3$ ), 1.18 (c, 1H,  $J = 10.03 \text{ Hz}$ , 4-HH), 1.88–1.92 (m, 2H, 4-HH + 3-H), 2.09 (dt, 1H,  $J = 7.52, 9.93 \text{ Hz}$ , 1-H), 3.22 (dd, 1H,  $J = 5.86, 7.47 \text{ Hz}$ , CHHN), 3.24–3.30 (m, 1H, CHHN), 3.33 (t, 1H,  $J = 7.62 \text{ Hz}$ , CHHOH), 3.37 (dd, 1H,  $J = 5.29, 7.62 \text{ Hz}$ , CHHOH), 4.74 (t, 1H,  $J = 5.00 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., NH), 7.70 (s, 1H, 2H<sub>pyrimidin</sub>).  $^{13}\text{C}$  NMR (DMSO- $d_6$ ):  $\delta = 16.46$  ( $\text{CH}_3$ ), 25.10 ( $\text{CH}_3$ ), 31.43 ( $\text{CH}_2$ ), 39.06 (C), 40.85 ( $\text{CH}_2$ ), 42.26 (CH), 44.15 (CH), 61.95 ( $\text{CH}_2$ ), 123.77 (CH), 136.85 (C), 145.97 (C), 152.26 (C). Anal. calcd. for  $\text{C}_{12}\text{H}_{19}\text{ClN}_4\text{O}$ : C, 53.23; H, 7.07; N, 20.69. Found: C, 53.31; H, 7.05; N, 20.58.

**(+)-(1*R*,cis)-3-(6-Chloro-9H-purin-9-yl)metil-2,2-dimethylcyclobutylmethanol (3a).** A mixture of **2a** (1.50 g, 5.54 mmol), triethyl orthoformate (30 mL), 12N HCl (1.4 mL) was stirred at room temperature for 24 h and then concentrated under vacuum obtain a residue that was treated with 0.5N HCl (30 mL) and THF (15 mL) for 2 h at room temperature. The resulting solution was brought to pH 7 with 1N NaOH, and evaporation of the solvents under reduced pressure left a residue that was chromatographed (silica gel  $\text{CH}_2\text{Cl}_2/\text{MeOH}$ , 20:1) afforded **3a** as a yellow oil (0.72 g, 47%).  $[\alpha]_{\text{D}}^{25} + 22.24$  (c 0.24, MeOH). IR (film):  $\nu = 3343, 2959, 1596, 1558, 1506, 1400, 1211, 648 \text{ cm}^{-1}$ .  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta = 0.94$  (s, 3H,  $\text{CH}_3$ ), 1.03 (s, 3H,  $\text{CH}_3$ ), 1.38 (c, 1H,  $J = 10.32 \text{ Hz}$ , 4-HH), 1.73 (dt, 1H,

$J = 7.86, 10.55 \text{ Hz}$ , 4-HH), 1.87 (ddt, 1H,  $J = 7.18, 8.10, 10.85 \text{ Hz}$ , 1-H), 2.36–2.45 (m, 1H, 3-H), 3.28 (dd, 1H,  $J = 5.89, 10.85 \text{ Hz}$ , CHHOH), 3.39 (ddd, 1H,  $J = 5.38, 8.21, 10.85 \text{ Hz}$ , CHHOH), 4.12 (dd, 1H,  $J = 8.75, 13.81 \text{ Hz}$ , CHHN), 4.23 (t, 1H,  $J = 5.02 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., OH), 4.31 (dd, 1H,  $J = 7.03, 13.81 \text{ Hz}$ , CHHN), 8.69 (s, 1H, 2-H<sub>purine</sub>), 8.75 (s, 1H, 8-H<sub>purine</sub>).  $^{13}\text{C}$  NMR (DMSO- $d_6$ ):  $\delta = 16.50$  ( $\text{CH}_3$ ), 24.61 ( $\text{CH}_3$ ), 30.76 ( $\text{CH}_2$ ), 39.22 (C), 41.16 ( $\text{CH}_2$ ), 43.85 (CH), 44.85 (CH), 61.64 ( $\text{CH}_2$ ), 131.19 (CH), 147.72 (C), 149.32 (CH), 151.72 (C), 152.18 (C). Anal. calcd. for  $\text{C}_{13}\text{H}_{17}\text{ClN}_4\text{O}$ : C, 55.62; H, 6.10; N, 19.96. Found: C, 55.31; H, 6.25; N, 19.83.

**(+)-(1*S*,cis)-6,9-Dihydro-9-[3-(hydroxymethyl)-2,2-dimethylcyclobutylmethyl]-1*H*-purin-6-ona (4a).** A mixture of **3a** (0.40 g) and 0.25N NaOH (17 mL) was refluxed for 6 h., the solvents under reduced pressure left a solid residue that was chromatographed (silica gel AcOEt/MeOH 7:3) afforded **4a** as a solid (0.34 g, 89%). Mp 198–200 °C (MeOH).  $[\alpha]_{\text{D}}^{25} + 98.33$  (c 0.3, MeOH). IR (KBr):  $\nu = 3445, 2958, 1648, 1586, 1548, 1521, 1457, 1364, 1021 \text{ cm}^{-1}$ .  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta = 0.94$  (s, 3H,  $\text{CH}_3$ ), 1.01 (s, 3H,  $\text{CH}_3$ ), 1.34 (c, 1H,  $J = 10.33 \text{ Hz}$ , 4-HH), 1.72 (dt, 1H,  $J = 7.86, 10.55 \text{ Hz}$ , 4-HH), 1.89 (ddt, 1H,  $J = 6.22, 8.03, 10.14 \text{ Hz}$ , 3-H), 2.44 (tt, 1H,  $J = 7.85, 8.98 \text{ Hz}$ , 1-H), 3.25 (dd, 1H,  $J = 6.45, 10.97 \text{ Hz}$ , CHHOH), 3.37 (dd, 1H,  $J = 8.10, 10.97 \text{ Hz}$ , CHHOH), 3.96 (dd, 1H,  $J = 8.60, 13.73 \text{ Hz}$ , CHHN), 4.15 (dd, 1H,  $J = 7.19, 13.73 \text{ Hz}$ , CHHN), 4.23 (s, 1H,  $\text{D}_2\text{O}$  exch., OH), 8.01 (s, 1H, 2-H<sub>purine</sub>), 8.05 (1s, s, 1H, 8-H<sub>purine</sub>), 12.29 (s, 1H,  $\text{D}_2\text{O}$  exch., NH).  $^{13}\text{C}$  NMR (DMSO- $d_6$ ):  $\delta = 16.50$  ( $\text{CH}_3$ ), 24.62 ( $\text{CH}_3$ ), 30.83 ( $\text{CH}_2$ ), 39.19 (C), 41.57 ( $\text{CH}_2$ ), 42.86 (CH), 43.31 (CH), 61.68 ( $\text{CH}_2$ ), 124.25 (CH), 140.55 (C), 145.73 (CH), 148.64 (C), 157.07 (C). Anal. calcd. for  $\text{C}_{13}\text{H}_{18}\text{N}_4\text{O}_2$ : C, 59.53; H, 6.92; N, 21.36. Found: C, 59.31; H, 6.65; N, 21.33.

**(+)-(1*S*,cis)-6,7-Dihydro-3-[3-(hydroxymethyl)-2,2-dimethylcyclobutyl]-3*H*-1,2,3-triazolo[4,5-*d*]pyrimidin-7-one (5a).** A solution of  $\text{NaNO}_2$  (0.12 g, 1.61 mmol) in water (3 mL) was added dropwise to a mixture of **2a** (0.35 g, 1.30 mmol), and 1N HCl (7 mL) at approx.  $-5^\circ\text{C}$ , the mixture was stirred at room temperature for 12 h, and removal of the solvents left a solid (0.32 g) that was chromatographed (silica gel AcOEt) afforded **5a** as a white solid (0.31 mg, 93%). Mp 184–186 °C (Cyclohexane/AcOEt/MeOH).  $[\alpha]_{\text{D}}^{25} + 7.76$  (c 0.27, MeOH). IR (KBr):  $\nu = 3308, 3054, 1717, 1653, 1636, 1591, 1457, 1384, 1010 \text{ cm}^{-1}$ .  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta = 0.90$  (s, 3H,  $\text{CH}_3$ ), 1.05 (s, 3H,  $\text{CH}_3$ ), 1.39 (c, 1H,  $J = 10.39 \text{ Hz}$ , 4-HH), 1.80 (dt, 1H,  $J = 7.84, 10.44 \text{ Hz}$ , 4-HH), 1.93 (ddt, 1H,  $J = 6.20, 8.18, 9.92 \text{ Hz}$ , 3-H), 2.42 (tt, 1H,  $J = 7.76, 9.29 \text{ Hz}$ , 1-H), 3.16 (dd, 1H,  $J = 6.20, 11.17 \text{ Hz}$ , CHHOH), 3.38 (dd, 1H,  $J = 8.18, 11.17 \text{ Hz}$ , CHHOH), 4.03 (t, 1H,  $J = 5.02 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., OH), 4.23 (dd, 1H,  $J = 7.50, 13.99 \text{ Hz}$ , CHHN), 4.52 (dd, 1H,  $J = 8.02, 13.99 \text{ Hz}$ , CHHN), 7.88 (s, 1H, 2H<sub>pyrimidine</sub>), 12.57 (s, 1H,  $\text{D}_2\text{O}$  exch., NH).  $^{13}\text{C}$  NMR (DMSO  $d_6$ ):  $\delta = 16.42$  ( $\text{CH}_3$ ), 24.64 ( $\text{CH}_3$ ), 30.68 ( $\text{CH}_2$ ), 41.41 (C), 42.30 ( $\text{CH}_2$ ), 43.93 (CH), 47.53 (CH), 61.64 ( $\text{CH}_2$ ), 129.78 (CH), 148.69 (C), 149.85 (C), 155.75 (C). Anal. calcd. for  $\text{C}_{12}\text{H}_{17}\text{N}_5\text{O}_4$ : C, 54.74; H, 6.51; N, 26.60. Found: C, 54.71; H, 6.45; N, 26.44.

**(+)-(1*R*,cis)-3-(7-Amino-3*H*-1,2,3-triazolo[4,5-*d*]pyrimidin-3-ylmetil)-2,2-dimethylcyclobutylmetanol (6a).** To a solution of **2a** (0.35 g, 1.30 mmol) in 1N HCl (3.5 mL) in a salted ice bath (approx  $-5^\circ\text{C}$ ), a solution of  $\text{NaNO}_2$  (0.12 g, 1.61 mmol) in water (13 mL) was added slowly enough to prevent the temperature from rising above  $0^\circ\text{C}$ . The mixture was stirred in the salted ice bath for 30 min, treated with 14N  $\text{NH}_4\text{OH}$  (7 mL), and refluxed for 1 h. After removal of water by azeotropic distillation with toluene and EtOH, the resulting solid residue (0.36 g) was chromatographed (silica gel  $\text{CH}_2\text{Cl}_2/\text{MeOH}$ , 8:1 and 7:1). Removal of the solvents from the latter eluate under reduced pressure left a solid residue (0.30 g) that upon recrystallization from Hex/AcOEt/MeOH afforded **6a** (0.22 g, 65%) as a white solid. Mp 194–196 °C.  $[\alpha]_{\text{D}}^{25} + 6.20$  (c 0.25,

MeOH). IR (KBr):  $\nu = 3745, 3150, 2950, 1665, 1607, 1576, 1510, 1457, 1384, 730 \text{ cm}^{-1}$ .  $^1\text{H NMR}$  (DMSO- $d_6$ ):  $\delta = 0.87$  (s, 3H,  $\text{CH}_3$ ), 1.06 (s, 3H,  $\text{CH}_3$ ), 1.39 (c, 1H,  $J = 10.33 \text{ Hz}$ , 4-HH), 1.79 (dt, 1H,  $J = 7.83, 10.50 \text{ Hz}$ , 4-HH), 1.92 (ddt, 1H,  $J = 6.27, 7.96, 11.06 \text{ Hz}$ , 1-H), 2.45 (tt, 1H,  $J = 7.81, 10.24 \text{ Hz}$ , 3-H), 3.27 (dd, 1H,  $J = 6.27, 10.86 \text{ Hz}$ , CHHOH), 3.38 (dd, 1H,  $J = 8.10, 10.86 \text{ Hz}$ , CHHOH), 4.22 (t, 1H,  $J = 5.05 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., OH), 4.33 (dd, 1H,  $J = 7.41, 13.98 \text{ Hz}$ , CHHN), 4.53 (dd, 1H,  $J = 8.14, 13.98 \text{ Hz}$ , CHHN), 8.03 (s, 1H,  $\text{D}_2\text{O}$  exch., NHH), 8.28 (s, 1H, 2H<sub>pyrimidine</sub>), 8.33 (s, 1H,  $\text{D}_2\text{O}$  exch., NHH).  $^{13}\text{C NMR}$  (DMSO- $d_6$ ):  $\delta = 16.42$  ( $\text{CH}_3$ ), 24.65 ( $\text{CH}_3$ ), 30.67 ( $\text{CH}_2$ ), 41.30 (C), 42.30 ( $\text{CH}_2$ ), 43.94 (CH), 47.12 (CH), 61.67 ( $\text{CH}_2$ ), 124.06 (CH), 148.93 (C), 156.55 (C), 156.92 (C). Anal. calcd. for  $\text{C}_{12}\text{H}_{18}\text{N}_6\text{O}$ : C, 54.95; H, 6.92; N, 32.04. Found: C, 54.83; H, 6.75; N, 32.23.

**(1S,cis)-N-[3-(Hydroxymethyl)-2,2-dimethylcyclobutyl]-N'-(3-ethoxypropanoyl)urea (8b)**. Dry benzene (50 mL) was added in the dark, under argon, to silver cyanate (7.5 g) that had previously been dried in vacuo over  $\text{P}_2\text{O}_5$  at  $100^\circ\text{C}$ . The resulting suspension was refluxed with vigorous stirring for 30 min, after which a solution of 3-ethoxypropenoyl chloride (3.0 g, 25 mmol) in dry benzene (10 mL) was added dropwise. The resulting mixture was refluxed with vigorous stirring for a further 30 min at room temperature for 3 h, and then left to settle. A sample of the supernatant (2.5 mL, theoretically containing 1.03 mmol of 3-ethoxypropenoyl isocyanate) was transferred to a dropping funnel and added dropwise under argon to a solution of **7b**<sup>[44]</sup> (0.10 g, 0.77 mmol) in dry DMF (4 mL) at  $-15^\circ\text{C}$ . The resulting mixture was left for 1 h to reach room temperature and was then stirred overnight and concentrated under reduced pressure (oil pump) at a temperature below  $40^\circ\text{C}$ . Removal of the solvents by repeated co-evaporation with ethanol left a solid that was chromatographed (silica gel  $\text{CHCl}_3/\text{EtOH}$ , 95:5) affording compound **8b** (0.12 g, 61%) a white solid. An analytic sample was obtained by crystallization from cyclohexane/EtOAc. Mp  $139\text{--}140^\circ\text{C}$ . IR (KBr):  $\nu = 3281, 2985, 1673, 1547, 1234, 1176, 1120 \text{ cm}^{-1}$ .  $^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 1.02$ . (s, 3H,  $\text{CH}_3$ ), 1.21 (s, 3H,  $\text{CH}_3$ ), 1.35 (t, 3H,  $J = 7.05 \text{ Hz}$ ,  $\text{CH}_2\text{CH}_3$ ), 1.48–1.57 (m, 1H, 4-HH), 1.94–2.01 (m, 1H, 4-HH), 2.33 (dt, 1H,  $J = 7.82, 10.88 \text{ Hz}$ , 3-H), 3.47–3.63 (m, 3H,  $\text{CH}_2\text{OH} + \text{OH}$ ), 3.90–3.99 (m, 3H,  $\text{CH}_2\text{CH}_3 + 1\text{-H}$ ), 5.31 (d, 1H,  $J = 12.22 \text{ Hz}$ , CH-OEt), 7.65 (d, 1H,  $J = 12.82 \text{ Hz}$ , CHCO), 8.72 (d, 1H,  $J = 7.75 \text{ Hz}$ , NHCH), 9.17 (s, 1H, CONHCO).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 15.15$  ( $\text{CH}_3$ ), 16.30 ( $\text{CH}_3$ ), 16.82 ( $\text{CH}_3$ ), 36.61 ( $\text{CH}_2$ ), 43.24 (CH), 49.30 (C), 64.72 ( $\text{CH}_2$ ), 76.73 (CH), 98.75 (CH), 153.82 (CH), 162.33 (C), 170.18 (C). Anal. calcd. for  $\text{C}_{13}\text{H}_{22}\text{N}_2\text{O}_4$ : C, 57.76; H, 8.20; N, 10.36. Found: C, 57.65; H, 8.35; N, 10.63.

**(1S,cis)-1-[3-(Hydroxymethyl)-2,2-dimethylcyclobutyl]-1,2,3,4-tetrahydropyrimidine-2,4-dione (9b)**. To a solution of **8b** (0.18 g, 0.7 mmol) in dioxane (11 mL) was added 2N  $\text{H}_2\text{SO}_4$  (14 mL), and this mixture was refluxed for 30 min, allowed to cool, brought to pH 7 with 2N NaOH, and concentrated to dryness. The residue was extracted with ethanol ( $3 \times 30 \text{ mL}$ ), and concentration of the extracts left a residue that was chromatographed (silica gel EtOAc/MeOH, 5:1) affording **9b** (0.15 g, 83%). Mp  $135\text{--}138^\circ\text{C}$ .  $^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 0.75$ . (s, 3H,  $\text{CH}_3$ ), 1.16 (s, 3H,  $\text{CH}_3$ ), 1.60–1.99 (m, 1H, 4-HH), 2.03–2.09 (m, 2H, 4-HH + 3-H), 3.34–3.74 (m, 2H,  $\text{CH}_2\text{OH}$ ), 4.28 (t, 1H,  $J = 9.20 \text{ Hz}$ , 1-H), 5.45 (d, 1H,  $J = 7.90 \text{ Hz}$ , 4-H<sub>pyrimidine</sub>), 7.64 (d, 1H,  $J = 7.95 \text{ Hz}$ , 5-H<sub>pyrimidine</sub>).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 15.58$  ( $\text{CH}_3$ ), 24.20 ( $\text{CH}_3$ ), 30.09 ( $\text{CH}_2$ ), 41.05 (CH), 44.25 (C), 56.78 (CH), 61.51 ( $\text{CH}_2$ ), 100.26 (CH), 143.19 (CH), 152.41 (C), 164.42 (C). Anal. calcd. for  $\text{C}_{11}\text{H}_{16}\text{N}_2\text{O}_3$ : C, 58.91; H, 7.19; N, 12.49. Found: C, 58.99; H, 7.17; N, 12.35.

**(±)-cis-N-(2-Hydroxy-1-indanyl)-N'-(3-ethoxypropenoyl)urea (11c)**. To a solution of aminoalcohol **10c**<sup>[45]</sup> (500 mg, 3.06 mmol) in DMF (22 mL) under Ar at  $-25^\circ\text{C}$ , a solution of 3-Ethoxy-2-propenoyl Isocyanate 1.15 M in benzene (3.91 mL, theoretical isocyanate

content 630 mg, 4.46 mmol) was added slowly enough to cause no rise in temperature, following which the reaction mixture was stirred overnight at r.t.. Toluene and EtOH were then added to form a low-boiling ternary azeotrope that was evaporated under reduced pressure while the temperature was maintained below  $40^\circ\text{C}$ , affording **11c** (524 mg, 56%) as a white solid. Mp  $170\text{--}2^\circ\text{C}$ . IR (KBr):  $\nu = 3293, 1701, 1666, 1626, 1542, 1474, 1335, 1256, 1205, 1187, 965, 794, 761 \text{ cm}^{-1}$ .  $^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 1.21\text{--}1.27$  (t, 3H,  $J = 7.05 \text{ Hz}$ ,  $\text{CH}_3$ ), 1.46–1.57 (dt, 1H,  $J = 7.76, 12.39 \text{ Hz}$ , 2 $\beta$ -H), 2.40–2.51 (dt, 1H,  $J = 7.14, 12.39 \text{ Hz}$ , 2 $\alpha$ -H), 3.12–3.18 (t, 1H,  $J = 6.32 \text{ Hz}$ , 1 $\beta$ -H), 3.35–3.42 (part A of an ABM system,  $J_{AB} = 12.64 \text{ Hz}$ ,  $J_{AM} = 6.91 \text{ Hz}$ , CHHNH), 3.62–3.69 (part B of an ABM system,  $J_{BA} = 12.64 \text{ Hz}$ ,  $J_{BM} = 6.99 \text{ Hz}$ , CHHNH), 3.90–3.98 (q, 2H,  $J = 7.01 \text{ Hz}$ ,  $\text{CH}_2\text{CH}_2\text{O}$ ), 4.94–4.99 (m, 1H, 3 $\beta$ -H), 5.32–5.35 (d, 1H,  $J = 5.75 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., OH), 5.48–5.53 (d, 1H,  $J = 12.30 \text{ Hz}$ , COCH), 7.21–7.35 (m, 4H, ArH), 7.53–7.58 (d, 1H,  $J = 12.30 \text{ Hz}$ , CH<sub>2</sub>OEt), 8.65 (s, 1H,  $\text{D}_2\text{O}$  exch.,  $\text{CH}_2\text{NHCO}$ ), 10.07 (s, 1H,  $\text{D}_2\text{O}$  exch., CONHCO).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 16.2$  ( $\text{CH}_3$ ), 42.1 ( $\text{CH}_2$ ), 43.1 (CH), 44.4 ( $\text{CH}_2$ ), 69.1 ( $\text{CH}_2$ ), 74.5 (CH), 100.1 (CH), 125.1 (CH), 125.9 (CH), 128.6 (CH), 129.2 (CH), 144.7 (C), 148.7 (C), 155.7 (C), 163.9 (CH), 169.5 (C). Anal. calcd. for  $\text{C}_{16}\text{H}_{20}\text{N}_2\text{O}_4$ : C, 63.14; H, 6.62; N, 9.20. Found: C, 63.31; H, 6.45; N, 9.33.

**1-(1-Inden-3-ylmethyl)-1,2,3,4-tetrahydropyrimidine-2,4-dione (12c)**. A solution of **11c** (75 mg, 0.246 mmol) and 2N  $\text{H}_2\text{SO}_4$  (5 mL) in 1,4-dioxane (2 mL) was refluxed 30 min. The mixture was allowed to cool and the solvent was evaporated under reduced pressure, and the resulting solid residue was chromatographed (silica gel,  $\text{CH}_2\text{Cl}_2/\text{MeOH}$ , 95:5) affording **12c** (47 mg, 79%) as a white solid. Mp  $175\text{--}180^\circ\text{C}$ . IR (KBr):  $\nu = 3019, 1718, 1559, 1458, 1423, 1380, 1340, 1242, 1202, 830, 745, 716 \text{ cm}^{-1}$ .  $^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 3.37$  (d, 2H,  $J = 1.30 \text{ Hz}$ , 1'-H), 4.86 (d, 2H,  $J = 1.40 \text{ Hz}$ ,  $\text{CH}_2\text{N}$ ), 5.62 (d, 1H,  $J = 7.93 \text{ Hz}$ , 5-H), 6.41 (s, 1H, 2'-H), 7.14 (d, 1H,  $J = 7.93 \text{ Hz}$ , 6-H), 7.16–7.32 (m, 3H, ArH), 7.43 (d, 1H,  $J = 6.66 \text{ Hz}$ , ArH), 9.05 (bs, 1H,  $\text{D}_2\text{O}$  exch., NH).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ ):  $\delta = 37.8$  ( $\text{CH}_2$ ), 44.5 ( $\text{CH}_2$ ), 101.7 (CH), 119.4 (CH), 124.3 (CH), 125.4 (CH), 126.5 (CH), 131.8 (CH), 139.8 (C), 142.9 (C), 144.4 (C), 145.5 (CH), 151.3 (C), 164.0 (C). Anal. calcd. for  $\text{C}_{14}\text{H}_{12}\text{N}_2\text{O}_2$ : C, 69.99; H, 5.03; N, 11.66. Found: C, 69.82; H, 5.17; N, 11.80.

**(1S,2R)-cis-N-(2-Hydroxy-1-indanyl)-N'-(3-ethoxypropenoyl)urea (14d)**. To a solution of aminoalcohol **13d** (500 mg, 3.35 mmol) in DMF (22 mL) under Ar at  $-25^\circ\text{C}$ , a solution of 3-ethoxy-2-propenoyl isocyanate 1.15 M in benzene (3.91 mL, theoretical isocyanate content 630 mg, 4.46 mmol) was added slowly enough to cause no rise in temperature, following which the reaction mixture was stirred overnight at r.t.. Toluene and EtOH were then added to form a low-boiling ternary azeotrope that was evaporated under reduced pressure while the temperature was maintained below  $40^\circ\text{C}$ , affording **14d** (576 mg, 59%) as a white solid.  $[\alpha]_D^{24} + 24.18$  (c 0.5, DMSO). Mp  $180\text{--}2^\circ\text{C}$ . IR (KBr):  $\nu = 3648, 3421, 1697, 1675, 1611, 1559, 1540, 1458, 1177, 1128, 816, 744 \text{ cm}^{-1}$ .  $^1\text{H NMR}$  (DMSO- $d_6$ ):  $\delta = 1.21\text{--}1.27$  (t, 3H,  $J = 7.05 \text{ Hz}$ ,  $\text{CH}_3$ ), 1.54–1.65 (dt, 1H,  $J = 7.58, 12.52 \text{ Hz}$ , 2 $\beta$ -H), 2.78–2.89 (dt, 1H,  $J = 7.18, 12.52 \text{ Hz}$ , 2 $\alpha$ -H), 3.87–3.95 (q, 2H,  $J = 7.05 \text{ Hz}$ ,  $\text{CH}_2\text{CH}_2\text{O}$ ), 5.41–5.44 (d, 1H,  $J = 6.35 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., OH), 5.48–5.53 (d, 1H,  $J = 12.13 \text{ Hz}$ , COCH), 4.89–4.97 (q, 1H,  $J = 6.45 \text{ Hz}$ , 2 $\beta$ -H), 5.03–5.13 (q, 1H,  $J = 8.12 \text{ Hz}$ , 1 $\beta$ -H), 7.16–7.36 (m, 4H, ArH), 7.47–7.52 (d, 1H,  $J = 12.13 \text{ Hz}$ , CH<sub>2</sub>OEt), 8.84–8.87 (d, 1H,  $J = 8.12 \text{ Hz}$ ,  $\text{D}_2\text{O}$  exch., CHNHCO), 10.15 (s, 1H,  $\text{D}_2\text{O}$  exch., CONHCO).  $^{13}\text{C NMR}$  (DMSO- $d_6$ ):  $\delta = 14.7$  ( $\text{CH}_3$ ), 40.1 ( $\text{CH}_2$ ), 57.6 (CH), 67.6 ( $\text{CH}_2$ ), 72.1 (CH), 98.8 (CH), 124.2 (CH), 125.0 (CH), 126.7 (CH), 127.6 (CH), 142.52 (C), 145.9 (C), 153.7 (C), 162.3 (CH), 167.9 (C). Anal. calcd. for  $\text{C}_{15}\text{H}_{18}\text{N}_2\text{O}_4$ : C, 62.06; H, 6.25; N, 9.65. Found: C, 61.97; H, 6.39; N, 9.81.

**(1S,2R)-cis-1-(2-Hydroxy-1-indanyl)-1,2,3,4-tetrahydropyrimidine-2,4-dione (15d)**. A solution of **14d** (36 mg; 0.124 mmoles) and 2N

H<sub>2</sub>SO<sub>4</sub> (8 mL) in 1,4-dioxane (4 mL) was refluxed 30 min. The mixture was allowed to cool, neutralized with NaOH 2N and the solvent was evaporated under reduced pressure. The resulting solid residue was chromatographed (silica gel, CH<sub>2</sub>Cl<sub>2</sub>/MeOH, 30:1) affording **15d** (27 mg, 89%) as a white solid. [ $\alpha$ ]<sub>D</sub><sup>24</sup> = -100.26 (c 1.15, CHCl<sub>3</sub>). Mp 99–102 °C. IR (KBr):  $\nu$  = 3402, 1688, 1463, 1387, 1261, 1055, 813, 747, 563 cm<sup>-1</sup>. <sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta$  = 2.87–2.94 (d, 1H, *J* = 16.95 Hz, 3'-H), 3.08–3.17 (dd, 1H, *J* = 5.62, 16.95 Hz, 3'-H), 4.65 (s, 1H, D<sub>2</sub>O exch., OH), 4.93 (s, 1H, 2'-H), 5.22–5.25 (d, 1H, *J* = 7.99 Hz, 5-H), 5.95–5.97 (d, 1H, *J* = 5.02 Hz, 1'-H), 6.90–6.93 (d, 1H, *J* = 7.99 Hz, 6-H), 7.08–7.19 (m, 4H, ArH), 10.77 (bs, 1H, D<sub>2</sub>O exch., NH). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = 39.2 (CH<sub>2</sub>), 63.6 (CH), 71.7 (CH), 100.2 (CH), 124.3 (CH), 125.8 (CH), 127.3 (CH), 128.9 (CH), 136.4 (C), 141.9 (C), 144.8 (CH), 152.1 (C), 165.4 (C). Anal. calcd. for C<sub>13</sub>H<sub>12</sub>N<sub>2</sub>O<sub>3</sub>: C, 63.93; H, 4.95; N, 11.47. Found: C, 64.11; H, 4.81; N, 11.39.

(±)-*cis*-7-(*tert*-Butyldimethylsilyloxymethyl)-5-(2-amino-6-chloro-9H-purin-9-ylmethyl)-1,4-diphenyl-6,7-dihydro-5H-cyclopenta[d]pyridazine (**18e**) and (±)-5-(*tert*-butyldimethylsilyloxymethyl)-7-methylidene-1,4-diphenyl-6,7-dihydro-5H-cyclopenta[d]pyridazine (**17e**). A solution of 2-amino-6-chloropurine (0.33 g, 1.97 mmol), 60% NaH (78.74 mg, 1.97 mmol) and 18-crown-6 ether (0.30 g, 1.14 mmol) in dry DMF (30 mL) was stirred under argon at 55 °C for 1 h. A solution of **16e**<sup>[43]</sup> (0.60 g, 1.14 mmol) in dry DMF (25 mL) was added, and stirring at 55 °C was continued for a further 24 h, after which the reaction mixture was diluted with CH<sub>2</sub>Cl<sub>2</sub> (50 mL) and washed with H<sub>2</sub>O (4 × 60 mL). The organic phase was dried over Na<sub>2</sub>SO<sub>4</sub>, and removal of the solvent left a solid residue (0.40 g) that was chromatographed (silica gel, hexane/EtOAc, 1:1 and 1:2). The early fractions afforded **17e**<sup>[47]</sup> (0.29 g, 59%) as a white solid; the middle fractions gave unreacted **16e** (20 mg); and the late fractions provided **18e** (66 mg, 10%). Mp 132–133 °C. IR (KBr):  $\nu$  = 3323, 2927, 2854, 1613, 1561, 1517, 1462, 1382, 1254, 1102, 911, 836, 775, 699 cm<sup>-1</sup>. <sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta$  = -0.30 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.23 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.65 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>), 1.96–2.05 (m, 1H, 6-H), 2.53 (dt, 1H, *J* = 14.0, 10.0 Hz, 6-H), 3.28 (dd, *J* = 10.13, 4.78 Hz, 1H, 7-H), 3.48 (dd, 1H, *J* = 10.13, 2.60 Hz, 5-H), 3.89–4.04 (m, 3H, OCH<sub>2</sub>+NCHH), 4.38–4.43 (m, 1H, NCHH), 5.12 (s, 2H, D<sub>2</sub>O exch., NH<sub>2</sub>), 7.39–7.51 (m, 6H), 7.80–7.77 (m, 4H), 7.97 (s, 1H, 8-H<sub>purine</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = -5.7 (SiCH<sub>3</sub>), -5.6 (SiCH<sub>3</sub>), 18.1 (C(CH<sub>3</sub>)<sub>3</sub>), 25.7 (C(CH<sub>3</sub>)<sub>3</sub>), 30.9 (CH<sub>2</sub>), 43.3 (CH), 46.1 (CH), 46.9 (CH<sub>2</sub>), 63.4 (CH<sub>2</sub>), 125.1 (CH), 128.4 (CH), 128.5 (CH), 128.6 (CH), 128.7 (CH), 129.3 (CH), 129.4 (CH), 136.4 (C), 136.6 (C), 141.9 (CH), 142.4 (C), 142.7 (C), 151.1 (C), 153.7 (C), 157.4 (C), 157.8 (C), 158.9 (C), 162.44 (C). MS (EI, 70 eV): *m/z* (%) = 600 (M+2, 29), 599 (M+1, 23), 598 (M, 62), 430 (32), 429 (100), 413 (5), 315 (6). Anal. calcd. for C<sub>32</sub>H<sub>36</sub>ClN<sub>7</sub>O<sub>5</sub>: C, 64.25; H, 6.07; N, 16.39. Found: C, 64.56; H, 6.18; N, 16.65.

(±)-*cis*-7-(2-Amino-6-chloro-9H-purin-9-ylmethyl)-1,4-diphenyl-6,7-dihydro-5H-cyclopenta[d]pyridazin-5-yl]methanol (**19e**): A 1 M solution of TBAF in THF (0.29 mL, 0.29 mmol) was added dropwise to a solution of **18e** (88 mg, 0.15 mmol) in the same solvent (4 mL) that was stirring under Ar in an ice bath. The solution was allowed to reach r.t., and stirring was continued for a further 30 min. The solvent was removed under reduced pressure and the residue was chromatographed (silica gel, hexane/EtOAc, 1:1). Concentration of the nonvoid fractions to dryness afforded **19e** (70 mg, 98%) as a white solid. Mp 177–178 °C. IR (KBr):  $\nu$  = 3313, 3197, 2930, 1616, 1566, 1524, 1468, 1411, 1380, 1279, 1170, 1052, 912, 773, 698 cm<sup>-1</sup>. <sup>1</sup>H NMR (DMSO-d<sub>6</sub>):  $\delta$  = 1.98–2.02 (m, 1H, 6-H), 2.52–2.59 (m, 1H, 6-H), 3.09–3.13 (m, 1H), 3.29–3.31 (m, 1H), 3.36–3.49 (m, 1H), 3.86–4.04 (m, 2H), 4.53–4.55 (m, 1H), 4.78 (t, 1H, *J* = 4.93 Hz, D<sub>2</sub>O exch., OH), 6.77 (s, 2H, D<sub>2</sub>O exch., NH<sub>2</sub>), 7.30–7.36 (m, 1H), 7.39–7.43 (m, 2H), 7.52–7.63 (m, 3H), 7.67 (s, 1H, 8-H<sub>purine</sub>), 7.72

(d, 2H, *J* = 7.33 Hz), 7.83–7.85 (m, 2H). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = 29.8 (CH<sub>2</sub>), 41.7 (CH), 45.4 (CH), 46.7 (CH<sub>2</sub>), 61.2 (CH<sub>2</sub>), 123.2 (C), 128.0 (CH), 128.1 (CH), 128.2 (CH), 128.6 (CH), 129.0 (CH), 136.0 (C), 136.9 (C), 142.5 (CH), 142.9 (C), 143.5 (C), 149.1 (C), 153.7 (C), 157.0 (C), 157.3 (C), 159.4 (C). MS (EI, 70 eV): *m/z* (%) = 486 (M+2, 21), 485 (M+1, 18), 484 (M, 69), 413 (6), 316 (35), 315 (100), 289 (6), 288 (31), 285 (13), 283 (10). Anal. calcd. for C<sub>26</sub>H<sub>22</sub>ClN<sub>7</sub>O: C, 64.53; H, 4.58; N, 20.26. Found: C, 64.92; H, 6.36; N, 16.87.

(±)-*cis*-7-(*tert*-Butyldimethylsilyloxymethyl)-5-(2-amino-6-cyclopentylamine-9H-purin-9-ylmethyl)-1,4-diphenyl-6,7-dihydro-5H-cyclopenta[d]pyridazine (**20e**). A solution of **18e** (103 mg, 0.17 mmol) and cyclopentylamine (0.10 mL, 1.04 mmol) in dry EtOH (10 mL) was refluxed under Ar for 30 h. Removal of the solvent under reduced pressure left a solid residue that was chromatographed (silica gel, hexane/EtOAc, 1:3). Concentration of the non-void fractions to dryness afforded **20e** (95 mg, 86%) as a white solid. Mp 113–115 °C. IR (KBr):  $\nu$  = 3325, 2952, 2858, 1596, 1484, 1444, 1394, 1252, 1101, 1033, 922, 836, 772, 730, 698 cm<sup>-1</sup>. <sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta$  = -0.29 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), -0.21 (s, 3H, Si(CH<sub>3</sub>)<sub>2</sub>), 0.67 (s, 9H, C(CH<sub>3</sub>)<sub>3</sub>), 1.48–1.52 (m, 2H), 1.62–1.65 (m, 2H), 1.71–1.73 (m, 2H), 2.03–2.07 (m, 3H), 2.50 (dt, 1H, *J* = 14.11, 9.98 Hz, 6-H), 3.23 (dd, 1H, *J* = 10.06, 5.33 Hz, 7-H), 3.47 (dd, 1H, *J* = 10.06, 2.90 Hz, 5-H), 3.88–3.95 (m, 3H, HOCH<sub>2</sub>+NCHH), 4.45–4.48 (m, 1H, NCHH), 4.57 (s, 2H, D<sub>2</sub>O exch., NH<sub>2</sub>), 5.44 (d, 1H, *J* = 7.69 Hz, D<sub>2</sub>O exch., NH), 7.00 (s, 1H, 8-H<sub>purine</sub>), 7.43–7.53 (m, 6H), 7.79–7.82 (m, 2H), 7.90 (d, 2H, *J* = 7.12 Hz). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = -5.7 (CH<sub>3</sub>), -5.6 (CH<sub>3</sub>), 18.2 (C(CH<sub>3</sub>)<sub>3</sub>), 23.7 (CH<sub>2</sub>), 25.8 (C(CH<sub>3</sub>)<sub>3</sub>), 30.7 (CH<sub>2</sub>), 33.5 (CH<sub>2</sub>), 43.7 (CH), 46.1 (CH), 46.2 (CH<sub>2</sub>), 63.6 (CH<sub>2</sub>), 128.5 (CH), 128.6 (CH), 128.7 (CH), 129.3 (CH), 136.6 (C), 136.7 (CH), 136.9 (C), 142.8 (C), 142.9 (C), 154.8 (C), 157.6 (C), 157.8 (C), 159.8 (C). MS (EI, 70 eV): *m/z* (%) = 648 (M+2, 47), 647 (M+1, 100), 646 (M, 39), 430 (33), 429 (91), 339 (21), 319 (11), 309 (17), 308 (11), 307 (21), 297 (12), 289 (11), 287 (16), 283 (22), 279 (12), 219 (17), 213 (16), 199 (31), 197 (50). Anal. calcd. for C<sub>37</sub>H<sub>46</sub>N<sub>8</sub>O<sub>5</sub>Si: C, 68.70; H, 7.17; N, 17.32. Found: C, 69.06; H, 7.33; N, 17.49.

(±)-*cis*-7-(2-Amino-6-cyclopentylamine-9H-purin-9-ylmethyl)-1,4-diphenyl-6,7-dihydro-5H-cyclopenta[d]pyridazin-5-yl]methanol (**21e**): A 1 M solution of TBAF in THF (0.27 mL, 0.27 mmol) was added dropwise to a solution of **20e** (88 mg, 0.15 mmol) in the same solvent (4 mL) that was stirring under Ar in an ice bath. The solution was allowed to reach r.t., and stirring was continued for a further 6 h. The solvent was removed under reduced pressure and the residue was chromatographed (silica gel, CH<sub>2</sub>Cl<sub>2</sub>/MeOH, 97:3 and 95:5). Concentration of the nonvoid fractions to dryness afforded **21e** (68 mg, 94%) as a white solid. Mp 145–146 °C. IR (KBr):  $\nu$  = 3331, 2948, 2865, 1599, 1486, 1446, 1396, 1347, 1244, 1077, 1045, 908, 766, 699 cm<sup>-1</sup>. <sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta$  = 1.48–1.52 (m, 2H), 1.58–1.65 (m, 2H), 1.69–1.73 (m, 2H), 1.91–1.95 (m, 1H), 2.02–2.05 (m, 2H, one of them D<sub>2</sub>O exch., OH+6-H), 2.37 (dt, 1H, *J* = 14.48, 9.96 Hz, 6-H), 3.36 (dd, 1H, *J* = 12.21, 3.70 Hz, 5-H), 3.54–3.60 (m, 2H, 7-H+HOCHH), 3.89 (d, 1H, *J* = 9.38 Hz, HOCHH), 3.97–4.03 (m, 1H), 4.08–4.17 (m, 1H, NCHH), 4.40–4.51 (m, 1H, NCHH), 4.62 (s, 2H, D<sub>2</sub>O exch., NH<sub>2</sub>), 5.74 (s, 1H, D<sub>2</sub>O exchange, NH), 7.89–7.87 (m, 2H), 7.10 (s, 1H, 8-H<sub>purine</sub>), 7.47–7.61 (m, 6H), 7.77–7.79 (m, 2H). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta$  = 23.7 (CH<sub>2</sub>), 28.8 (CH<sub>2</sub>), 44.0 (CH<sub>2</sub>), 45.9 (CH), 46.8 (CH), 61.9 (CH<sub>2</sub>), 113.9 (C), 128.3 (CH), 128.4 (CH), 128.8 (CH), 129.0 (CH), 129.2 (CH), 129.6 (CH), 136.3 (CH), 136.9 (C), 142.4 (C), 143.5 (C), 155.1 (C), 157.4 (C), 158.6 (C), 159.7 (C). MS (EI, 70 eV): *m/z* (%) = 534 (M+2, 24), 533 (64), 532 (M, 26), 391 (26), 369 (10), 316 (10), 315 (40), 313 (10), 311 (10), 306 (19), 289 (14), 285 (12), 283 (13), 271 (7), 219 (18), 199 (11), 167 (14), 165 (11). Anal. calcd. for C<sub>31</sub>H<sub>32</sub>N<sub>8</sub>O: C, 69.90; H, 6.06; N, 21.04. Found: C, 70.25; H, 6.19; N, 20.93.

(1*R*,*cis*)-3-[2-Amino-6-(cyclopropylamino)-9H-purin-9-ylmethyl]-1,2,2-trimethylcyclopentylmethanol (**23f**). A solution of **22f**<sup>[47]</sup> (100 mg, 0.32 mmol) and cyclopropylamine (183 mg, 3.20 mmol) in EtOH (12 mL) was refluxed for 20 h. Removal of the solvent under reduced pressure left a solid from which was chromatography (silica gel, EtOAc/hexane, 8:2). Concentration of the non-void fractions to dryness afforded **23f** (85 mg, yield 80%) as a white solid. M.p. 202–204 °C.  $[\alpha]_D^{25} + 28$  (c 0.1, MeOH). IR (KBr):  $\nu = 3346, 3222, 2964, 1652, 1596, 1487, 1399, 1356, 1027 \text{ cm}^{-1}$ . <sup>1</sup>H NMR (CDCl<sub>3</sub>):  $\delta = 0.58\text{--}0.63$  (m, 2H, cyclopropyl), 0.82–0.88 (m, 2H, cyclopropyl), 0.92 (s, 3H, CH<sub>3</sub>), 0.99 (s, 3H, CH<sub>3</sub>), 1.0 (s, 3H, CH<sub>3</sub>), 1.43–1.52 (m, 2H), 1.56–1.72 (m, 2H), 1.89 (b s, 1H, D<sub>2</sub>O exch., OH), 2.38–2.43 (m, 1H), 2.96–3.01 (m, 1H, 1-H<sub>cyclopropyl</sub>), 3.47 and 3.61 (AB system, 2H,  $J = 10.7 \text{ Hz}$ , CH<sub>2</sub>OH), 4.09 and 3.83 (AB part of a ABX system, 2H,  $J_{AB} = 13.5 \text{ Hz}$ ,  $J_{AX} = 10.3 \text{ Hz}$ ,  $J_{BX} = 4.5 \text{ Hz}$ , CH<sub>2</sub>N), 5.81 (b s, 1H, D<sub>2</sub>O exch., NH), 4.78 (b s, 2H, D<sub>2</sub>O exch., NH<sub>2</sub>), 7.46 (s, 1H, 8-H<sub>purine</sub>). <sup>13</sup>C NMR (CDCl<sub>3</sub>):  $\delta = 160.28$  (C), 156.55 (C), 152.46 (C), 137.77 (CH), 115.02 (C), 69.56 (CH<sub>2</sub>), 49.05 (C), 48.75 (3CH<sub>3</sub>), 45.35 (CH<sub>2</sub>), 44.72 (C), 33.85 (CH<sub>2</sub>), 30.11 (CH<sub>2</sub>), 26.90 (CH<sub>2</sub>), 24.07 (CH), 23.81 (CH), 21.22 (CH), 18.71 (CH), 7.82 (CH<sub>2</sub>). MS (EI, 70 eV):  $m/z$  (%) = 345 (M+1, 21), 344 (96), 329 (86), 299 (18), 204 (29), 203 (46), 191 (37), 190 (31), 189 (44), 175 (100), 174 (22), 173 (46), 163 (22), 162 (25). Anal. Calcd. for: C<sub>18</sub>H<sub>28</sub>N<sub>6</sub>O: C, 62.76; H, 8.19; N, 24.40. Found: C, 63.02; H, 8.33; N, 24.52.

### 3.2 Antitumor Activity

All assays were carried out in flat-bottomed 96-well microtiter plates. To each well were added  $5 \times 10^4$  murine leukemia cells (L1210/0) and a given amount of the test compound. The cells were allowed to proliferate for 48 hours at 37 °C on a humidified, CO<sub>2</sub>-controlled atmosphere. The growth of the cells was linear during this 48 h incubation period. At the end of the incubation period, the cells were

counted in a coulter counter (Coulter Electronics Ltd, Harpenden Herts, England) and the number of dead cells was evaluated by staining with trypan blue. The IC<sub>50</sub> (50% inhibitory concentration) was defined as the compound concentration that inhibit cell proliferation by 50%, as compared to untreated control<sup>[26]</sup>.

## 4 Results and Discussion

### 4.1 Model Setup

Our training set contains a diverse set of nucleosides analogues, derived from purinic and pyrimidinic bases with diverse values of inhibitory effects in the proliferation of L1210/0 cancer cells. Our first goal was to establish a discriminant function based on the most relevant 3D molecular descriptors, after ascertaining them with an adequate selection method. This model (from now on denoted as Model-1) is given below together with the statistical parameters of the linear discriminant analysis (LDA).

Model-1

$$\begin{aligned}
 P = & 1.417 \cdot J3D - 0.225 \cdot ADDD + 78.975 \cdot MEcc \\
 & + 0.841 \cdot RDF120v - 1.529 \cdot Mor19u \cdot 0.599 \cdot Mor12m \\
 & + 0.250 \cdot Mor04e + 5.598 \cdot E3s - 0.051 \cdot Au \\
 & + 0.136 \cdot ITH - 83.187
 \end{aligned}
 \quad (1)$$

$$\begin{aligned}
 N = 278 \quad \rho = 25.27 \quad F(10,267) = 17.610 \\
 \rho < 10^{-5} \quad \lambda = 0.602 \quad D^2 = 2.819
 \end{aligned}$$

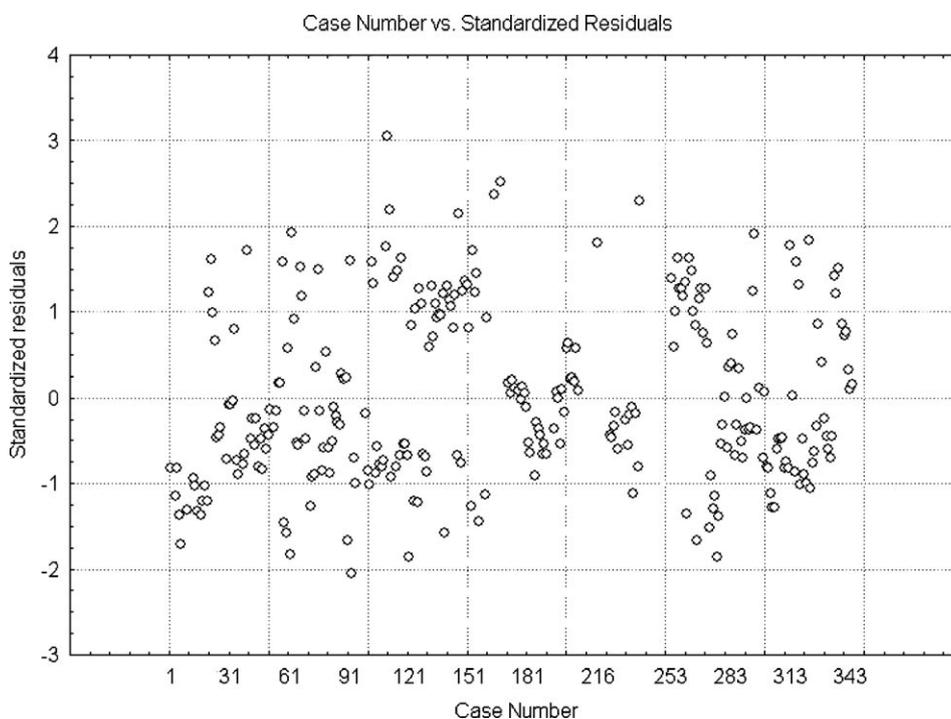


Figure 2. Distribution of the standardized residuals for all cases studied.

The  $\rho$  value, large sample size ( $N$ ), large  $F$  index, and small  $p$  value, are indicative of the model's statistical significance. In addition, the values of the Wilks'  $\lambda$  statistic ( $\lambda$  can take values from zero – perfect discrimination – to one – no discrimination –) and of the Mahalanobis distance (a measure of the separation between the active and inactive groups) show that the model displays an adequate discriminatory power for differentiating both groups. The latter is also confirmed by the classification results; the model correctly classified 81.4% (83 out of 102) of the 102 cytotoxic compounds and 81.8% (144 out of 176) of the 176 noncytotoxic compounds, giving rise to an overall 81.7% (227 out of 278) effective discrimination of the 278 training set compounds. Furthermore, the percentage of unclassified compounds in the training set is only 2.87% (8 out of 278), while the percentages of false inactives and false actives are 5.39% (15 out of 278) and 11.5% (32 out of 278), respectively.

Further analysis of this discriminant model should only be pursued after checking the reliability of the pre-adopted assumptions. Firstly, LDA establishes a linear additive relation between the molecular descriptors and the underlying bioactivity. In fact, this is the simplest mathematical form that might be envisaged for the model in the absence of any a priori information. Nevertheless, by looking at the distribution of the standardized residuals (observed minus predicted divided by the square root of the residual mean square) for all cases (Figure 2), no specific pattern can be seen, therefore suggesting that the model does not exhibit a non-linear dependence.

The assumption of non-multicollinearity between the descriptors can be readily confirmed by analyzing the correlation matrix. The results in Table 1 show that the pairs of 3D-descriptors (Au; ADDD), (Au; RDF10v), (ITH; ADDD), (ITH; RDF120v), and (ITH; Au) are highly correlated with each other, denoting thereby redundancy in the information displayed by such pairs. Therefore, we have examined the performance of orthogonal complements in modeling the anticancer activity.

Following Randić's procedure,<sup>[39]</sup> we determined the orthogonal complements for all variables in Model-1 that, after standardization, allowed us to derive the following best ten-variable equation (Model-2).

$$P = 1.357 \cdot \Omega^1 \text{Mor12m} + 0.806 \cdot \Omega^2 \text{ITH} - 0.958 \cdot \Omega^3 \text{ADDD} + 0.498 \cdot \Omega^4 \text{MEcc} + 0.455 \cdot \Omega^5 \text{Mor04e} + 0.452 \cdot \Omega^6 \text{E3s} - 0.400 \cdot \Omega^7 \text{Mor19u} - 0.334 \cdot \Omega^8 \text{Au} + 0.521 \cdot \Omega^9 \text{RDF120v} + 0.345 \cdot \Omega^{10} \text{J3D} - 0.303 \quad (2)$$

$$N = 278 \quad \rho = 25.27 \quad F(10, 267) = 17.244$$

$$p < 10^{-5} \quad \lambda = 0.607 \quad D^2 = 2.760$$

where the symbol  $\Omega^i X$  means the orthogonal complement of variable  $X$ , while the subscript refers to the order selected for orthogonalizing the variables. Table 2 summarizes and defines the descriptors used in this model.

As can be seen, all descriptors were found to be statistically significant and the overall fitness of the model preserved as the statistics are as robust as before (see Equation 1), though the classification of active compounds slightly improved (82.3%) with respect to that of the non-orthogonal model (81.4%). By comparing Equation 2 with Equation 1, one can see that there are no changes in either the sign of the regression coefficients or of the constant. Nevertheless the relative contributions of the variables in Equation 2 are significantly different from those in Equation 1. Therefore, for purposes of QSAR interpretability, we shall use the orthogonal-standardized descriptor model defined in Equation 2.

Another important parametric assumption of LDA is multivariate normality.<sup>[40]</sup> Figure 3 shows the plots for the histograms of frequency distribution of the discriminant function, divided by active and non-active groups, respectively, which takes into account the interaction among variables. Attached to each plot are also the results derived from the Kolmogorov–Smirnov statistical test ( $d$ ). Accordingly, a visual inspection of the normality plots and frequency distributions for the discriminant function (Figure 3), as well as the calculated  $d$  values for both groups (actives:  $d=0.099$ ; inactives:  $d=0.041$ ;  $p>0.200$  in both cases), lead us to accept the hypothesis of multivariate normality.

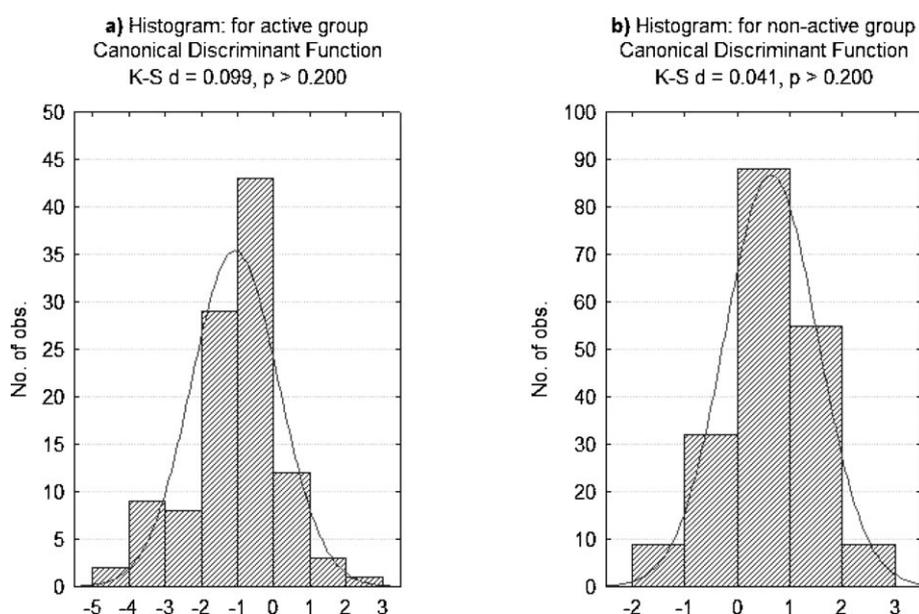
Let us now check the hypothesis of *homocedasticity*. A possible problem regarding the homogeneity of the (co)variances is suggested by the Box's  $M$  statistical test ( $p<0.01$ ), though this test can be overly sensitive to large data files<sup>[40]</sup>

**Table 1.** Intercorrelation between the ten descriptors used in QSAR model-1 (Eq. 1). Significant correlations are marked in bold.

	J3D	ADDD	MEcc	RDF120v	Mor19u	Mor12m	Mor04e	E3s	Au	ITH
J3D	1.00	-0.34	-0.26	-0.38	0.16	0.00	0.10	0.11	-0.39	-0.38
ADDD		1.00	0.05	0.70	0.74	-0.07	0.33	0.04	<b>0.84</b>	<b>0.88</b>
MEcc			1.00	0.08	-0.04	-0.19	0.13	-0.49	0.01	-0.18
RDF120v				1.00	0.45	0.04	0.14	0.07	<b>0.83</b>	<b>0.71</b>
Mor19u					1.00	-0.18	0.41	0.06	0.52	0.56
Mor12m						1.00	-0.33	0.32	0.16	0.24
Mor04e							1.00	-0.16	0.09	0.13
E3s								1.00	0.16	0.27
Au									1.00	<b>0.88</b>
ITH										1.00

**Table 2.** Symbols for the descriptors used in the QSAR models and their definition.

Symbol	Descriptor type	Descriptor definition
Mor12m	3D-MoRSE	3D-MoRSE – signal 12/weighted by atomic masses
Mor04e	3D-MoRSE	3D-MoRSE – signal 04/weighted by atomic Sanderson electronegativities
Mor19u	3D-MoRSE	3D-MoRSE – signal 19/unweighted
RDF120v	RDF	Radial Distribution Function – 12.0/weighted by atomic van der Waals volumes
ITH	GETAWAY	Total information content on the leverage equality
ADDD	Geometrical	Average distance/distance degree
MEcc	Geometrical	Molecular eccentricity
J3D	Geometrical	3D-Balaban index
E3s	WHIM	3 <sup>rd</sup> component accessibility directional WHIM index/weighted by atomic electrotopological states
Au	WHIM	A total size index/unweighted

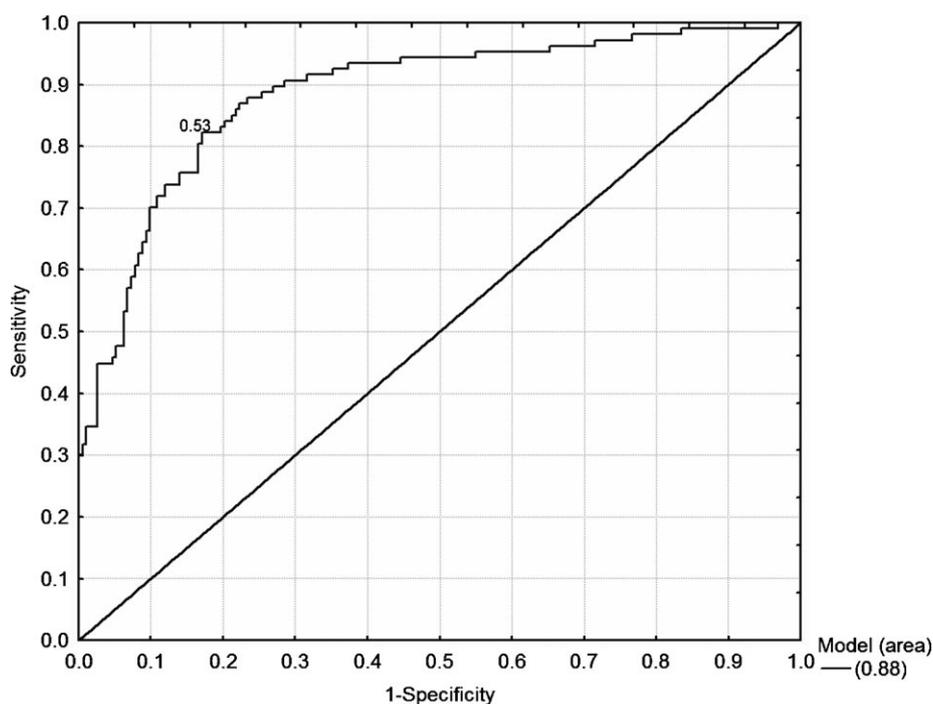
**Figure 3.** Histograms for the frequency distributions of the discriminant function (Eq. 1), considering active groups (a) and non-active groups (b).

which is likely what happened here. This nevertheless increases the likelihood that a case belongs to the higher dispersion group and, in this sense, adjusting the a priori probabilities can greatly improve the overall classification rate of the discriminant model.

A distinct, better threshold for the a priori classification probability can be estimated by means of the Receiver Operating Characteristics (ROC) curve.<sup>[41]</sup> This is a useful technique not only for obtaining the best thresholds but also for organizing classifiers. As seen in Figure 4, the optimal threshold for predicting the active chemicals with the present QSAR model is 0.53. Further, one can see that the model is not a random, but a truly statistically significant, classifier, since the area under the ROC curve (= 0.88) is significantly higher than the area under the random classifier curve (diagonal line).

Having checked the pre-adopted assumptions, it is now important to assess the model quality regarding its robustness and how well it might be expected to generalize, i.e. how well it will correctly predict the activity of new compounds. This was first accomplished here by means of internal cross-validation (CV) of the model, using the leave-group-out procedure. The statistics and classification results reported in Table 3 correspond to five independent leave-20%-out CV runs, each involving a different, randomly chosen partition into a training and a test set.

As can be seen, the model is robust and shows little dependence on the composition of the training and test sets, since it displays good statistics ( $\lambda$ ,  $D^2$  and  $F$ ) as previously (see Eq. 2). It also appears to show a good predictive power, judging from the averages computed for the overall classification results in the training and predictive sets generated in each CV run (81.4% and 79.8%, respectively).



**Figure 4.** Receiver operating characteristic (ROC) curve for the final classification model (Eq. 2).

**Table 3.** Results from the cross-validation leave-group-out procedure. Results obtained with Model-2 (Eq. 2) after removing ca.20% of compounds from the training set. %AC<sub>G</sub>(T) and %AC<sub>G</sub>(P) are the percentage of good overall classifications in the training and predicting sets, respectively.

CV-run	$\lambda$	$D^2$	$F$	%AC <sub>G</sub> (T)	%AC <sub>G</sub> (P)
1	0.615	2.675	13.164	81.0	82.7
2	0.624	2.581	12.733	80.2	85.5
3	0.593	2.927	14.557	81.6	80.0
4	0.596	2.886	14.352	82.1	78.2
5	0.585	3.021	15.026	82.1	72.7
average	0.603	2.818	13.966	81.4	79.8

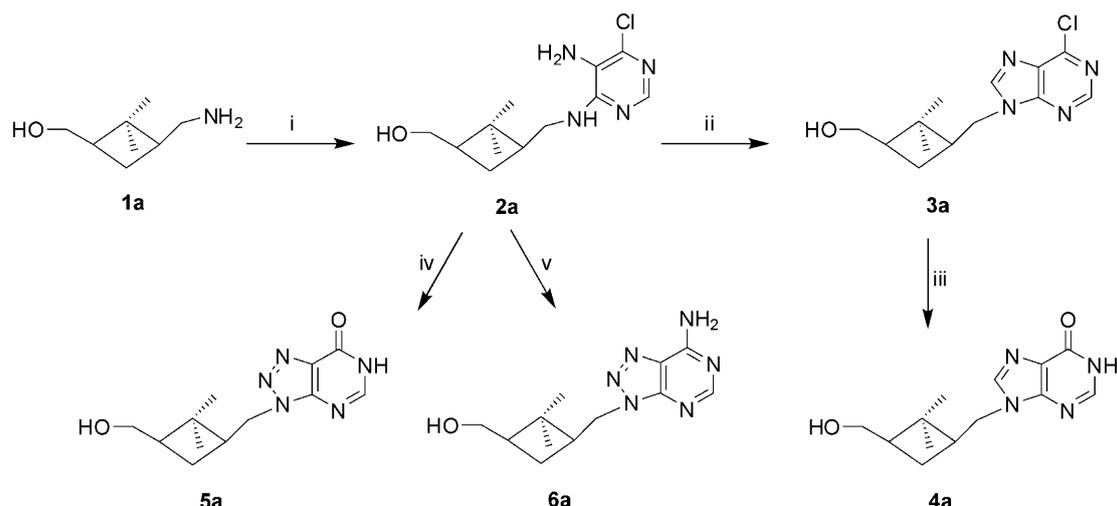
#### 4.2 Discovery of Novel Carbonucleoside Anticancer Compounds

In this work, a design approach for the rational selection (discovery) of novel compounds with anticancer activity is applied. The design is conceived here as “a preliminary outline showing the main features of something to be executed”. In this way, we choose 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 laboratories and others not even synthesized. The compounds so designed were evaluated by our discriminant model and when needed, synthesized afterwards (9 compounds). Notice that, even though some of these compounds were recognized by the model as inactive, they were nevertheless synthesized for the purpose of QSAR-model validation.

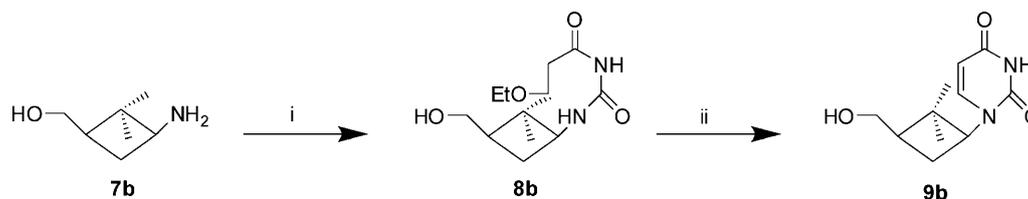
The first group of compounds (**4a**, **5a** and **6a**) was synthesized according to the strategy given in Scheme 1. As shown in this Scheme, compound **4a** was prepared using standard chemistry for purine carbocyclic nucleosides analogues: amino alcohol **1a**<sup>[42]</sup> was condensed with 5-amino-4,6-dichloropyrimidine, the resulting diamine **2a** was cyclized with ethyl orthoformate to obtain the intermediate **3a** (in 90% yield) and this 9-substituted 6-chloropurine was further transformed into the final inosine **4a** through hydrolysis with NaOH. The diazotation of compound **2a** under the usual conditions and with the usual subsequent procedures<sup>[43]</sup> gave the 8-azapurine derivatives **5a** and **6a** (see Scheme 1).

On the other hand, the second group of compounds (the uridine analogues **9b**, **12c** and **15d**) were constructed on the aminoalcohol **7b**,<sup>[44]</sup> **10c**<sup>[45]</sup> and **13d** by condensation with 3-ethoxy-2-propenoyl isocyanate<sup>[46]</sup> and cyclization of



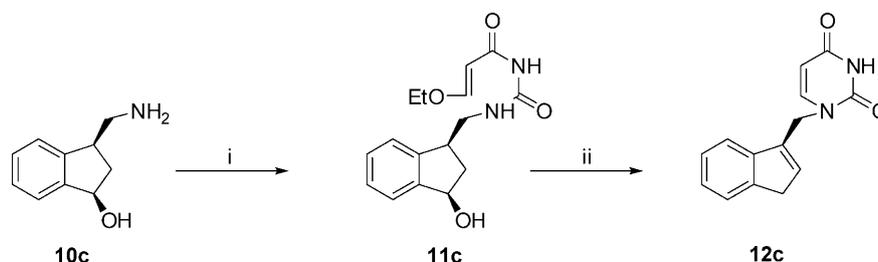
Reagents and conditions: i) 5-Amino-4,6-dichloropyrimidine, Et<sub>3</sub>N, *n*-BuOH, reflux, 72 h; ii) Triethyl orthoformate, 12N HCl, r.t. 24 h; iii) 0.25N NaOH, reflux, 6 h; iv) NaNO<sub>2</sub>, H<sub>2</sub>O, 1N HCl, r.t. 12 h; v) 1N HCl, NaNO<sub>2</sub>, H<sub>2</sub>O, 0°C, and 14N NH<sub>4</sub>OH, reflux, 1 h.

**Scheme 1.** General procedure for the preparation of the test set nucleoside analogues **3a**, **4a** and **5a**.



Reagents and conditions: i) 3-ethoxypropenoyl isocyanate, 15°C; ii) 2N H<sub>2</sub>SO<sub>4</sub>, reflux 30 min.

**Scheme 2.** General procedure for the preparation of the test set nucleoside analogue **9b**.



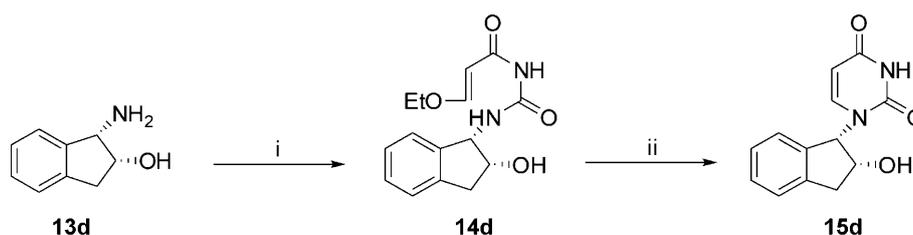
Reagents and conditions: i) Ethoxy-2-propenoyl isocyanate, DMF, 12 h, rt. ii) 2N H<sub>2</sub>SO<sub>4</sub>, dioxane, 30 min, reflux.

**Scheme 3.** General procedure for the preparation of the test set nucleoside analogue **12c**.

the urea **8b** with conventional acidic conditions (see Schemes 2, 3 and 4).

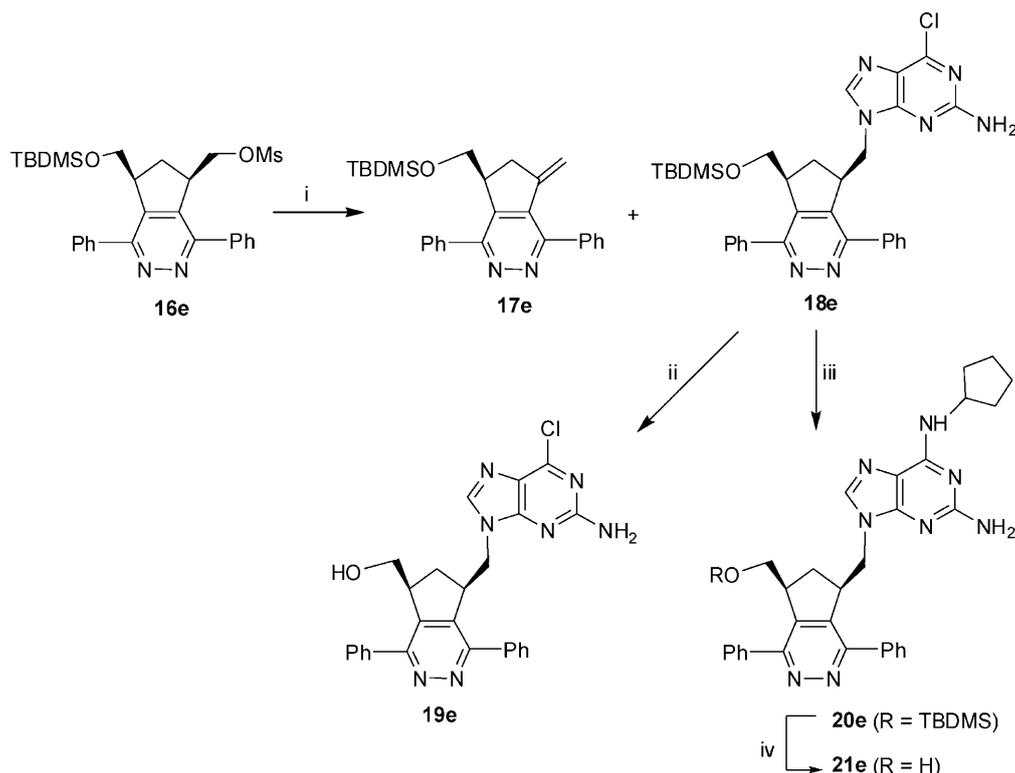
In a previous work, our research group synthesized 1(*N*)-homocarbanucleosides in which the double bond of the cyclopentene ring of carbovir and abacavir was replaced with a pyridazine.<sup>[10]</sup> Following that work, we have then synthesized the third group of compounds (**19e**, **21e**) that bear the 2-amino group of carbovir and abacavir. Such com-

pounds were prepared from known mesylate **16e**<sup>[47]</sup>: heating crude **16e** for 24 h at 55 °C with 2-amino-6-chloropurine, NaH and 18-crown-6 ether (see Scheme 5) afforded a 10% yield of **18e** (lower temperatures and shorter reaction times gave even smaller yields) together with unreacted **16e** (10%) and a 59% yield of **17e**,<sup>[47]</sup> the product of a competing elimination reaction that was shown by TLC monitoring to produce detectable amounts of **17e** within



*Reagents and conditions:* i) Ethoxy-2-propenoyl isocyanate, DMF, 12 h, rt. ii) 2N H<sub>2</sub>SO<sub>4</sub>, dioxane, 30 min, reflux.

**Scheme 4.** General procedure for the preparation of the test set nucleoside analogue **15d**.



*Reagents and conditions:* i) 2-amino-6-chloropurine, 60% NaH, 18-crown-6, DMF, 55°C (1 h); ii) 1M TBAF in THF, r.t. (0.5 h); iii) cyclopentylamine, EtOH, reflux (30 h); iv) 1M TBAF in THF, r.t. (6 h).

**Scheme 5.** General procedure for the preparation of the test set nucleoside analogues **19e** and **21e**.

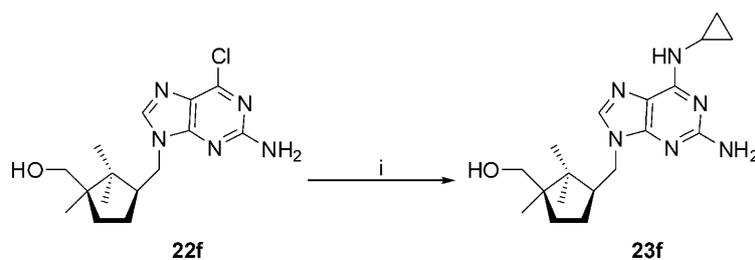
just a few minutes of the start of the reaction. Refluxing **18e** with cyclopentylamine in ethanol gave **20e**. Finally, deprotection of **18e** and **20e** with TBAF afforded the new purinyl carbonucleosides **19e** and **21e**.

Finally, the reaction of (1*R*,*cis*)-3-(6-chloro-9*H*-purin-9-ylmethyl)-1,2,2-trimethylcyclopentylmethanol (**22f**)<sup>[47]</sup> with cyclopentylamine afforded the new purinyl carbonucleoside **23f**.

The carbonucleoside analogues now synthesized and the others previously synthesized comprise a new set (22 compounds) for external validation, which in turn is an absolute requirement towards a final, truly assessment of the predic-

tive power of our QSAR model. So, following on, an experimental evaluation of the inhibitory effects of these compounds on the proliferation of L1210/0 cancer cells was carried out. Four of them have already been evaluated experimentally,<sup>[47]</sup> but the remaining have not. The results of the classification and experimental pharmacological evaluations are given in Table 4.

As can be seen in Table 4, Model-2 (Eq. 2) correctly classifies 80% (4 out of 5) of actives NAs and 88% (15 out of 17) of inactives. At the same time, the percentage of overall classification is 86% (19 out of 22), corroborating the good predictive ability of the present discriminant model. It can



Reagents and conditions: i) cyclopropylamine, EtOH, reflux (20 h).

**Scheme 6.** General procedure for the preparation of the test set nucleoside analogue **23f**.

**Table 4.** The 22 carbonucleosides used in the external prediction set along with the observed cytotoxicity against the cellular line L1210/0 and the classification (a posteriori probabilities) according to QSAR model-2 (Eq. 2).

No.	$IC_{50}$ ( $\mu$ M) <sup>[a]</sup>	prob (%) <sup>[b]</sup>	Predicted class <sup>[c]</sup>	Reference
<b>Active chemicals</b>				
24h	7.89	99.236	+1	[10]
25g	42.46	95.449	+1	[12]
21e	65.71	94.772	+1	this work
26g	176.76	90.270	+1	[12]
23f	13.94	5.019	-1	this work
<b>Non-active chemicals</b>				
27j	> 200	97.451	-1	[21]
28j	> 200	96.722	-1	[21]
29j	> 200	96.671	-1	[21]
5a	> 200	96.545	-1	this work
30j	> 200	96.457	-1	[21]
4a	> 200	95.572	-1	this work
6a	> 200	94.772	-1	this work
9b	> 200	92.981	-1	this work
12c	> 200	91.788	-1	this work
31k	> 200	88.022	-1	[68]
32k	> 200	86.458	-1	[68]
33i	> 200	78.445	-1	[11]
34k	> 200	74.102	-1	[68]
15d	> 200	73.074	-1	this work
35i	> 200	59.836	-1	[11]
36g	> 200	43.843	+1	[12]
19e	> 200	18.807	+1	this work

[a] 50% inhibitory concentration or compound concentration required to reduce the proliferation of tumors cells by 50%. [b] A posteriori probability of classifying a chemical as active or inactive, according to Equation 2. [c] Values of +1 and -1 stand for compounds with and without cytotoxic activity, respectively.

be also seen that, of the six compounds that were theoretically classified by our model as active ones, four turn out to be really inhibitors, and were thus considered as well-

classified. The other two compounds resulted inactive in the experiment evaluation and should then be considered as false actives. Furthermore, compound **26g** showed some

activity against the cellular line; however, the values of  $IC_{50}$  for this compound are very high (the highest of all active compounds) revealing that this compound has a poor antitumor activity.

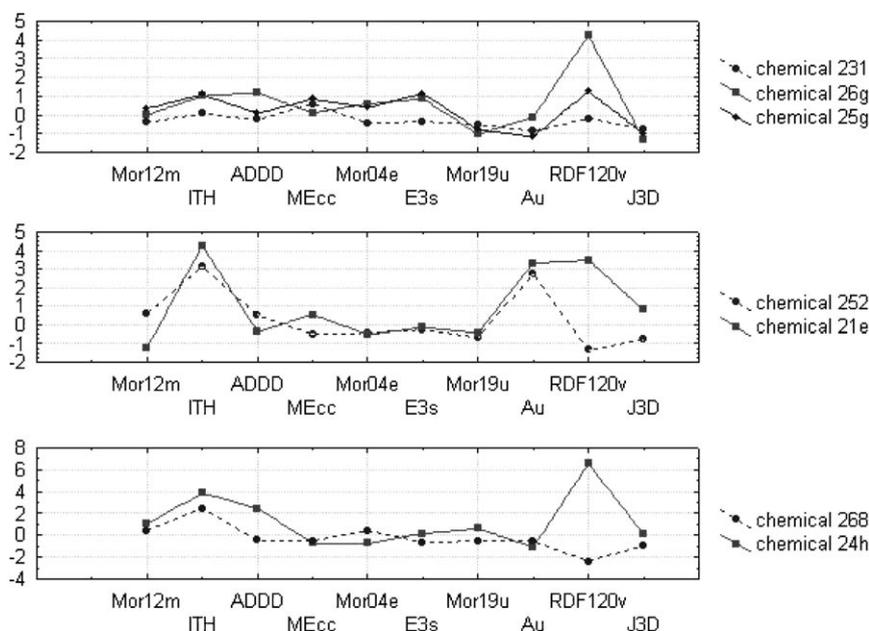
The two false active compounds constitute 9% of bad classification of the entire data set of carbonucleosides evaluated. Of these, compounds **36g** and **19e** are similar to compounds **26g** and **21e**, respectively. In the first case, the only difference between them is the change of the NH-cyclopropyl group to a  $C_6H_5CH_3$  group (**36g** vs. **26g**), while in the second case the difference is the change of the Cl atom to a NH-cyclopentyl group (**19e** vs. **21e**). From the experimental values of the antitumor activity, it appears that these differences are enough for suppressing that activity. But nevertheless these structural differences are not well-captured by the present approach.

Overall, five carbonucleoside compounds (**13f**, **21e**, **24h**, **25g**, **26g**) of the external set resulted to be active ( $IC_{50} < 200 \mu M$ ) in the biological assays,<sup>[26]</sup> of which four were well predicted by eq 2. The latter were formerly designed by a careful interpretation of QSAR Model-2. In this model, ten descriptors are correlated with the anticancer activity (see Eq. 2 and Table 2), but the most important variables for these four compounds are the descriptors<sup>[34]</sup>: RDF12v – obtained by the Radial Distribution Function approach using a spherical distance of 12 Å weighted by atomic van der Waals volumes; Mor04e and Mor12m – calculated by summing atomic weights (atomic electronegativities and masses, respectively) viewed by a different electron diffraction angular scattering function; ITH – calculated as the total information content provided by the leverage matrix; and E3s – calculated by the projection onto the principal components obtained from the weighted (in this case, by

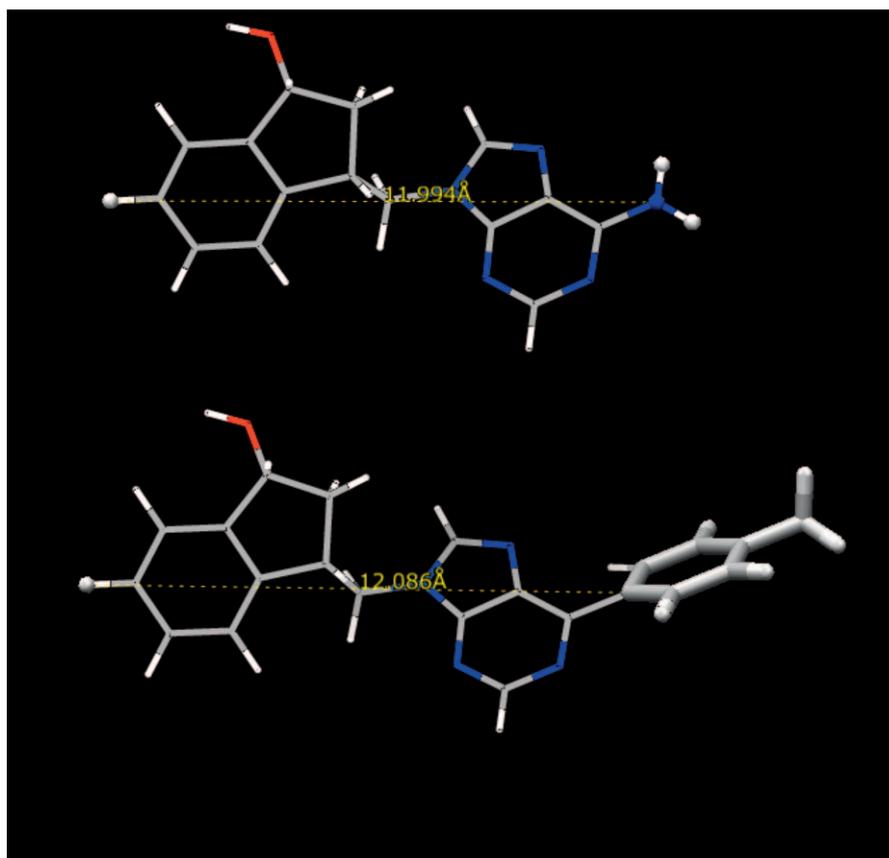
atomic electrotopological states) covariance matrix of atomic coordinates. A plot of the molecular descriptor values for each of these compounds is shown in Figure 5.

Based on this interpretation of Model-2, we thus tried to modify the structural features of the noncytotoxic carbonucleosides of the training set in order to turn them into cytotoxic ones. For instance, Figure 6 shows a 3D-structural sketch of the inactive chemical **231** and of the new active chemicals **26g** (for simplicity, chemical **25g** is not shown as it has a similar pattern to chemical **26g**). As can be seen, both NAs have the same carbocycles, but the substituents on the  $C^6$ -position of the purine base differ. For improving the anticancer activity, we try to increase the atomic van der Waals volume of the substituent at a distance of ca. 12 Å, which is around the  $C^6$ -position of purine (see Figure 5). In so doing, by replacing its target NH2 substituent by a *p*-chlorophenyl (chemical **25g**) or by a *p*-tolyl (chemical **26g**), that really enhanced the corresponding RDF120v values of the new designed chemicals (see Figures 5 and 6). In addition, the descriptors Mor04e, E3s and ITH also rose for both chemicals (see Figure 6). These descriptors codify electronic factors (Mor04e, E3s) and molecular complexity (ITH). In the particular case of the ITH descriptor, its discriminant power is associated to its strong dependence on molecular size.<sup>[48]</sup>

A similar approach was followed on for the inactive chemical **252**, which allowed us to design the new active chemical **21e** (see Figure 7). Both are derivatives of the biphenylcyclopentylpyridazine, but their substituents on the purine base differ. By increasing the bulkiness of the substituent at the  $C^6$ -purine position, e.g.: from cyclopropyl to cyclopentyl (**252** vs. **21e**), as well as adding an amino group on the  $C^4$ -purine position (compound **21e**), we could



**Figure 5.** Values of molecular descriptors Mor12m, ITH, ADD, MEcc, Mor04e, E3s, Mor19u, Au, RDF120v and J3D (eq 2) for the inactive chemicals (dashed line) and active chemicals (solid line).



**Figure 6.** 3D structures of the inactive-chemical **231** (top) and active-chemical **26g** (bottom). The maximum interatomic distances to the substituent on the purine C<sup>6</sup>-position are shown. Substructures represented by sticks correspond to distances greater than 12 Å.

then increase the value of the RDF120v descriptor (see Figure 7), and so the biological activity. As can be seen, Mor12m is also a discriminant descriptor between both compounds but unlike descriptor RDF120v, its value decreased after the structural modification. As according to the model (Eq. 2) it makes a positive contribution to the biological activity, it seems that the descriptor Mor12m has a less effect on the anticancer activity of these structures than the RDF120v descriptor.

Finally, by examining the structural details of chemicals **268** (inactive) and **24h** (active), it can be seen that the replacement of the hydroxymethyl group located on the C<sup>2</sup>-position of the cyclopentyl moiety by the tert-butyldimethylsilyl-*O*-methoxyl group did improve the biological activity. By inspection of Figure 5, one can notice a significant increase of the RDF120v value when going from the inactive chemical (**268**) to the new active one (**24h**). It can be easily understood by looking at Figure 8, which shows that the degree of bulkiness of the substituent located on the C<sup>2</sup>-position of the cyclopentyl ring, at an interatomic distance greater than 12 Å, is greater for the active carbonucleoside than for the inactive one. In other words, a sterically bulky substituent at the C<sup>2</sup>-position of the cyclopentyl

ring condensate to pyridazine is favorable to improve the anticancer activity. This was also noticed in the previous cases (see Figures 6 and 7), but in other positions of the molecular structure. The remaining descriptors did not show a significant variation.

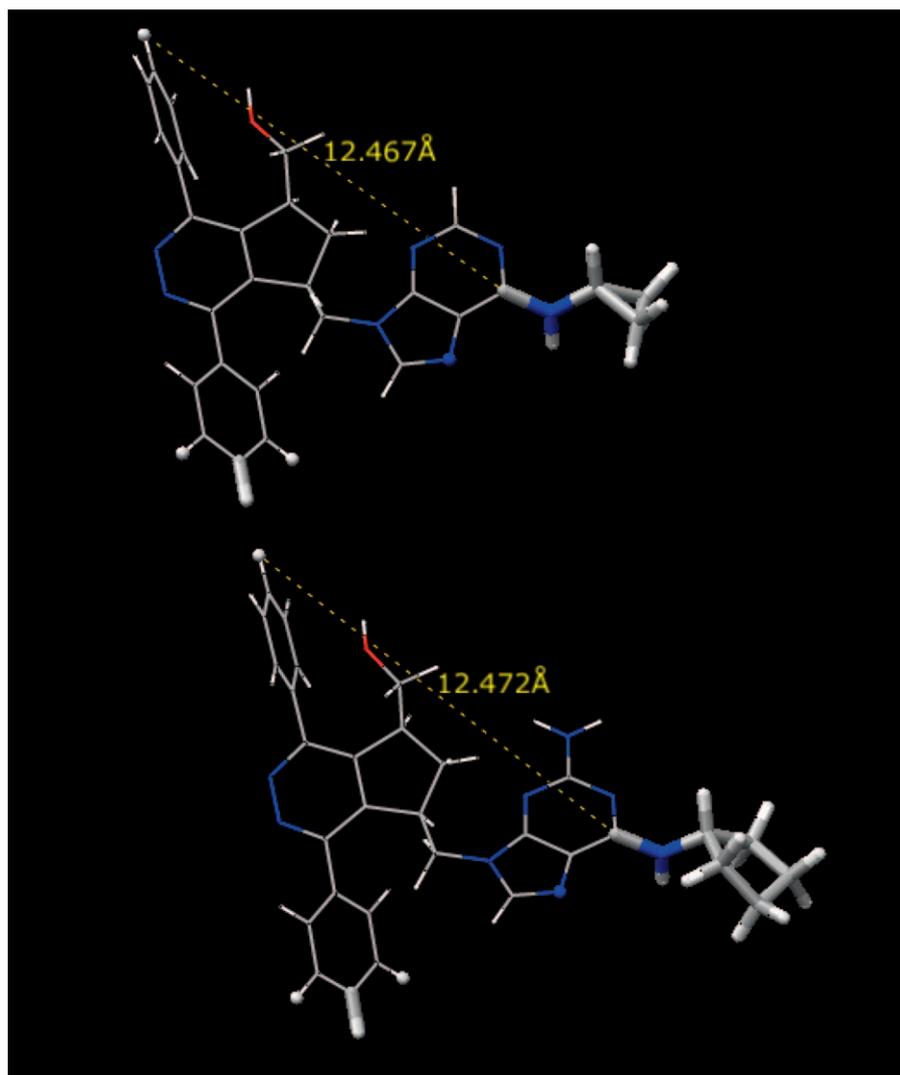
### 3.3 2D and 3D Modeling in Balance

In a previous work, we developed a discriminant function based on 2D-DRAGON descriptors.<sup>[33]</sup> The following 2D-QSAR model was derived from a training set of 241 NA compounds, by combining LDA along with the same variable selection technique.

$$P = 1.828 \cdot {}^2\Omega D / Dr06 + 3.646 \cdot {}^3\Omega piPC10 + 0.455 \cdot {}^5\Omega MATS8e - 3.708 \cdot {}^6\Omega piPC09 - 0.549 \cdot {}^7\Omega MPC10 + 0.161 \quad (3)$$

$$N = 241 \quad \rho = 48.2 \quad F(7,233) = 35.74 \quad p < 10^{-5}$$

$$\lambda = 0.57 \quad D^2 = 3.21$$

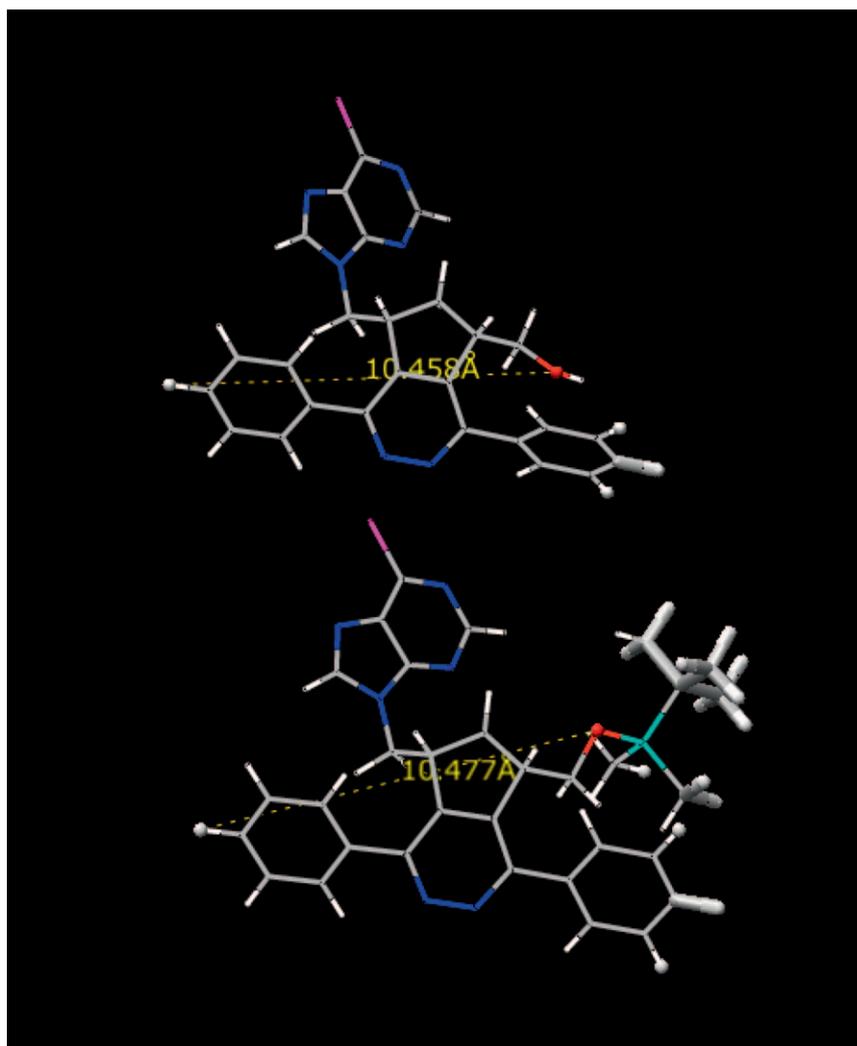


**Figure 7.** 3D structures of the inactive-chemical **252** (top) and active-chemical **21e** (bottom). The maximum interatomic distances to the substituent on the purine C<sup>6</sup>-position are shown. Substructures represented by sticks correspond to distances greater than 12 Å.

Though 2D-DRAGON molecular descriptors can encode information about adjacency, branching, etc., they cannot take into account information regarding conformational aspects. With the desire to build a reliable QSAR model based on 3D-DRAGON descriptors, which could be used to flag even better potential anticancer activity, we recognized a unique opportunity when we could assemble a set of over 270 NAs tested in a single, consistent cytotoxic assay.

For a comparison with the 3D-QSAR model built here, our former 2D-QSAR model<sup>[33]</sup> (see Eq. 3) was also used for predicting the anticancer activity of the present external test set. In so doing, we found that the model correctly classified 80% (4 out of 5) of the active compounds, like the 3D-QSAR model, and 94% (16 out of 17) of the inactives ones, thereby revealing a slightly superior percentage of external overall classification. Yet, the predicted data can only be considered reliable for those compounds that fall

within the applicability domain on which the model was derived.<sup>[49]</sup> Figure 9 shows a Williams plot, i.e. the plot of the standardized residuals (y-axis) versus the computed leverage values (x-axis) for each compound of the training set. From this plot, one can establish the applicability domain as the squared area within  $\pm 2$  standard deviations and the leverage threshold  $h^*$  ( $h^*$  is generally fixed at  $3p'/n$ , where  $n$  is the number of training compounds and  $p'$  the number of model parameters). The leverage thresholds for the 2D and 3D models are  $h^*=0.074$  and  $h^*=0.119$ , respectively. As can be noticed, for the 2D-QSAR model, 5 out of 22 compounds of the external test set are outside of this area, and 3 of them have large standardized residual values (greater than 2 standard deviation units). In contrast, for the 3D-QSAR model, only 2 out of the 22 carbocyclic nucleosides have a leverage greater than  $h^*$  and they show standard deviation values within the limits. Therefore the



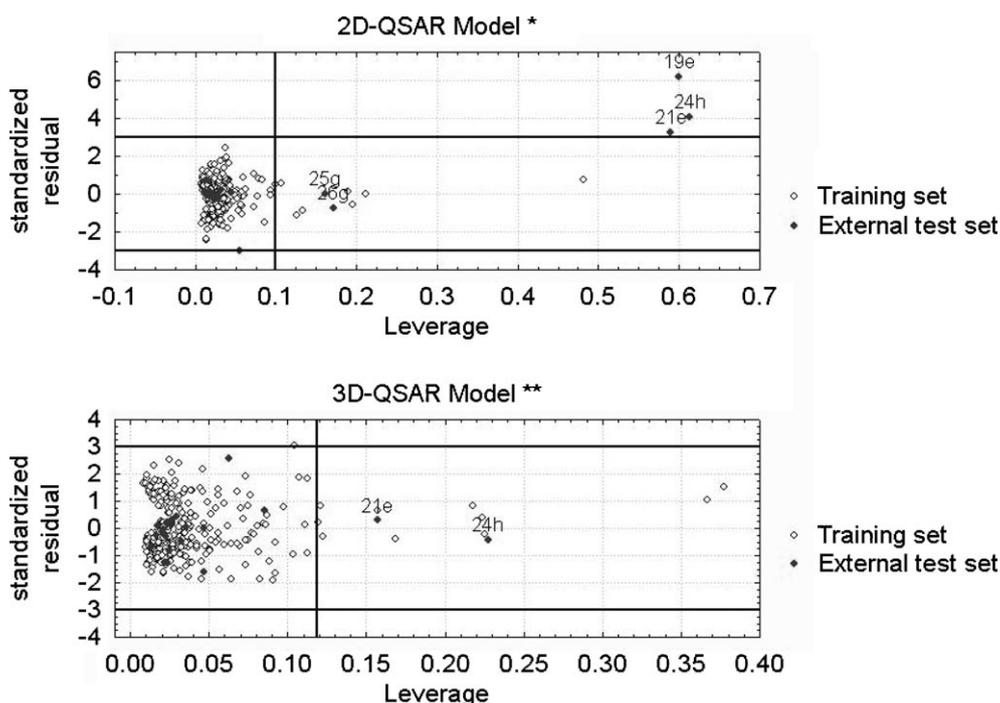
**Figure 8.** 3D structures of the inactive-chemical **268** (top) and active-chemical **24h** (bottom). The maximum interatomic distances to the oxygen atom of the hydroxymethyl group and the tert-butyldimethylsilyl-*O*-methoxyl group located on the C<sup>2</sup>-position of the cyclopentyl moiety are shown. Substructures represented by sticks correspond to distances greater than 12 Å.

3D-QSAR model derived here has a greater applicability domain than the 2D-QSAR model previously developed by us,<sup>[33]</sup> most likely due to the fact that it includes 3D-structural information and was derived using a larger training set.

Here it is important to remark that, if a chemical is outside the applicability domain according to a given, correctly applied method, this is not a final argument for rejecting the prediction; rather, it is an indication of the increase uncertainty of the prediction. We can say that this is, in a statistical sense, an incorrect application of a model, but it is nevertheless possible that the model will generate a correct result.

## 5 Conclusions

In this work, we have examined the ability of a large, diverse and consistently tested training set to provide predictive QSAR models for probing potential anticancer activity, specifically antileukemia activity. The training set included 278 NAs derived from purinic and pyrimidinic bases, and was assembled from literature compounds with published cytotoxic activity against L1210/0 cancer cells. For that purpose, we have thoroughly evaluated LDA models in conjunction with a variety of 3D structure representations and feature selection algorithms. For this training set, the best found 3D model showed good accuracy, robustness and predictivity, as judged by several statistics and extensive internal cross-validation. The final 3D-model showed also



**Figure 9.** Comparison between the applicability domains of our final 3D-QSAR model (Eq. 2) and of the previous 2D-QSAR model (Eq. 3).

good predictive power with an external test set of 22 NA compounds, which were previously designed by us. From this external set, nine compounds have never been synthesized, and the cytotoxic activity of eighteen of them is being reported here for the first time. Therefore, the model allowed the design, synthesis and biological evaluation of novel NAs, which is of great relevance as new compounds enrich the structural diversity of related databases and could be used in forthcoming QSARs. Although previous efforts with a 2D-classification model were successful, we recognized that 3D-QSAR models could give more useful information, primarily because they highlight the 3D structural features and stereochemistry of the compounds that are relevant to their biological activity. In addition, the 3D-QSAR model derived here was shown to have a greater applicability domain than that previously developed 2D-model. Therefore, the structural information gathered and the 3D-QSAR model per se will certainly help in the future design of novel potent anticancer drugs.

## Acknowledgment

The authors acknowledge the *Portuguese Fundação para a Ciência e a Tecnologia (FCT)* (SFRH/BD/22692/2005) and the *Xunta de Galicia* (Project 07CSA008203PR) for financial support, and appreciate the assistance of *Yunierkis Pérez-Castillo* for help in performing some of the calculations.

## References

- [1] a) H. Frankish, *Lancet* **2003**, 361, 1278; b) A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, M. J. Thun, *CA Cancer J. Clin.* **2008**, 58, 71.
- [2] C. Simons, *Carbocyclic Nucleosides*; Gordon and Breach, UK **2001**.
- [3] a) J. M. Blanco, O. Caamaño, F. Fernandez, X. García-Mera, A. R. Hergueta, C. Lopez, J. E. Rodriguez-Borges, J. Balzarini, E. De Clercq, *Chem. Pharm. Bull. (Tokyo)* **1999**, 47, 1314; b) M. J. Figueira, J. M. Blanco, O. Caamaño, F. Fernández, X. García-Mera, C. López, *Arch. Pharm. Pharm. Med. Chem.* **1999**, 332.
- [4] M. T. Crimmins, *Tetrahedron* **1998**, 54, 9229.
- [5] M. I. Nieto, O. Caamaño, F. Fernández, G. Gómez, J. Balzarini, E. De Clercq, *Nucleosides, Nucleotides Nucleic Acids* **2002**, 21, 243.
- [6] M. J. Figueira, O. Caamaño, F. Fernández, J. E. Rodríguez-Borges, J. Balzarini, E. De Clercq, *Synthesis* **2004**, 12, 1991.
- [7] A. R. Hergueta, F. Fernandez, C. Lopez, J. Balzarini, E. De Clercq, *Chem. Pharm. Bull. (Tokyo)* **2001**, 49, 1174.
- [8] C. Balo, F. Fernández, E. Lens, C. López, *Arch. Pharm. Pharm. Med. Chem.* **1997**, 265.
- [9] J. M. Blanco, O. Caamaño, F. Fernandez, J. E. Rodriguez-Borges, J. Balzarini, E. de Clercq, *Chem. Pharm. Bull. (Tokyo)* **2003**, 51, 1060.
- [10] S. F. Wnuk, C. S. Yuan, R. T. Borchardt, J. Balzarini, E. De Clercq, M. J. Robins, *J. Med. Chem.* **1997**, 40, 1608.
- [11] E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana, E. De Clercq, *J. Med. Chem.* **2000**, 43, 1975.
- [12] H. Gonzalez-Diaz, D. Viña, L. Santana, E. de Clercq, E. Uriarte, *Bioorg. Med. Chem.* **2006**, 14, 1095.
- [13] S. Prekupec, D. Svedruzic, T. Gazivoda, D. Mrvos-Sermek, A. Nagl, M. Grdisa, K. Pavelic, J. Balzarini, E. De Clercq, G. Folkers,

- L. Scapozza, M. Mintas, S. Raic-Malic, *J. Med. Chem.* **2003**, *46*, 5763.
- [14] A. A. Moosavi-Movahedi, S. Hakimelahi, J. Chamani, G. A. Khodarahmi, F. Hassanzadeh, F. T. Luo, T. W. Ly, K. S. Shia, C. F. Yen, M. L. Jain, R. Kulatheeswaran, C. Xue, M. Pasdar, G. H. Hakimelahi, *Bioorg. Med. Chem.* **2003**, *11*, 4303.
- [15] S. Raic-Malic, D. Svedruzic, T. Gazivoda, A. Marunovic, A. Hergold-Brundic, A. Nagl, J. Balzarini, E. De Clercq, M. Mintas, *J. Med. Chem.* **2000**, *43*, 4806.
- [16] L. Santana, M. Teijeira, E. Uriarte, J. Balzarini, E. De Clercq, *Eur. J. Med. Chem.* **2002**, *37*, 755.
- [17] G. H. Hakimelahi, N. W. Mei, A. A. Moosavi-Movahedi, H. Davari, S. Hakimelahi, K. Y. King, J. R. Hwu, Y. S. Wen, *J. Med. Chem.* **2001**, *44*, 1749.
- [18] G. H. Hakimelahi, A. A. Moosavi-Movahedi, T. Sambaiah, J. L. Zhu, K. S. Ethiraj, M. Pasdar, S. Hakimelahi, *Eur. J. Med. Chem.* **2002**, *37*, 207.
- [19] S. Hatse, L. Naesens, E. De Clercq, J. Balzarini, *Biochem. Pharmacol.* **1999**, *58*, 311.
- [20] C. Teran, L. Santana, M. Teijeira, E. Uriarte, E. De Clercq, *Chem. Pharm. Bull. (Tokyo)* **2000**, *48*, 293.
- [21] Z. X. Wang, L. I. Wiebe, J. Balzarini, E. De Clercq, E. E. Knaus, *J. Org. Chem.* **2000**, *65*, 9214.
- [22] S. F. Wnuk, C. S. Yuan, R. T. Borchardt, J. Balzarini, E. De Clercq, M. J. Robins, *J. Med. Chem.* **1994**, *37*, 3579.
- [23] N. G. Kundu, P. Das, J. Balzarini, E. De Clercq, *Bioorg. Med. Chem.* **1997**, *5*, 2011.
- [24] I. A. Mikhailopulo, N. E. Poopeiko, T. I. Pricota, G. G. Sivets, E. I. Kvasnyuk, J. Balzarini, E. De Clercq, *J. Med. Chem.* **1991**, *34*, 2195.
- [25] S. Raic-Malic, A. Hergold-Brundic, A. Nagl, M. Grdisa, K. Pavelic, E. De Clercq, M. Mintas, *J. Med. Chem.* **1999**, *42*, 2673.
- [26] E. De Clercq, J. Balzarini, P. F. Torrence, M. P. Mertes, C. L. Schmidt, D. Shugar, P. J. Barr, A. S. Jones, G. Verhelst, R. T. Walker, *Mol. Pharmacol.* **1981**, *19*, 321.
- [27] E. De Clercq, J. Descamps, J. Balzarini, J. Giziewicz, P. J. Barr, M. J. Robins, *J. Med. Chem.* **1983**, *26*, 661.
- [28] M. J. Robins, P. W. Hatfield, J. Balzarini, E. De Clercq, *J. Med. Chem.* **1984**, *27*, 1486.
- [29] R. N. Hunston, A. S. Jones, C. McGuigan, R. T. Walker, J. Balzarini, E. De Clercq, *J. Med. Chem.* **1984**, *27*, 440.
- [30] L. A. Al-Razzak, D. Schwepler, C. J. Decedue, J. Balzarini, E. De Clercq, M. P. Mertes, *J. Med. Chem.* **1987**, *30*, 409.
- [31] M. Bobek, S. H. An, D. Skrinicosky, E. De Clercq, R. J. Bernacki, *J. Med. Chem.* **1989**, *32*, 799.
- [32] E. De Clercq, J. Balzarini, J. Descamps, C. F. Bigge, C. T. Chang, P. Kalaritis, M. P. Mertes, *Biochem. Pharmacol.* **1981**, *30*, 495.
- [33] A. M. Helguera, J. E. Rodriguez-Borges, X. García-Mera, F. Fernandez, M. N. Cordeiro, *J. Med. Chem.* **2007**, *50*, 1537.
- [34] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, **2000**.
- [35] O. Caamaño, G. Gómez, F. Fernández, M. D. García, X. García-Mera, E. De Clercq, *Synthesis* **2004**, *17*, 2855.
- [36] J. M. Blanco, O. Caamaño, F. Fernández, G. Gómez, M. I. Nieto, J. Balzarini, E. Padalko, E. De Clercq, *Nucleosides Nucleotides* **1997**, *16*, 159.
- [37] a) F. Fernández, X. García-Mera, M. Morales, L. Vilariño, O. Caamaño, E. De Clercq, *Tetrahedron* **2004**, *60*, 9245; b) F. Fernández, X. García-Mera, C. López, M. Morales, J. E. Rodríguez-Borges, *Synthesis* **2005**, *20*, 3549; c) C. Balo, J. M. Blanco, F. Fernandez, E. Lens, C. López, *Tetrahedron* **1998**, *54*, 2833.
- [38] a) R. Todeschini, V. Consonni, M. Pavan, *Dragon Software*, Version 2.1. **2002**; b) J. Frank, *MOPAC*, Version 6.0, S. R. Laboratories, US Air Force Academy, Colorado Springs, CO, **1993**; c) *STATISTICA* (data analysis software system), Version 7.0, Statsoft Inc, **2002**.
- [39] M. Randić, *New. J. Chem.* **1991**, *15*, 517.
- [40] M. H. Pestana, J. N. Gageiro, *Análise de dados para ciências sociais. A complementaridade do SPSS*; Lisboa: Edições Sílabo, **2008**.
- [41] F. Provost, T. Fawcett, *Analysis and Visualization of Classifier Performance Comparison Under Class and Cost Distributions*, American Association for Artificial Intelligence, **1997**.
- [42] F. Fernández, C. López, A. R. Hergueta, *Tetrahedron* **1995**, *51*, 10317.
- [43] R. Vince, S. Daluge, *J. Org. Chem.* **1980**, *45*, 531.
- [44] C. Balo, O. Caamaño, F. Fernández, C. López, *Tetrahedron: Asymmetry* **2005**, *16*, 2593.
- [45] M. Morales, F. Fernández, X. García-Mera, J. E. Rodríguez-Borges, in *XI National Congress of the Spanish Society of Therapeutic Chemistry: Valencia (Spain)*, **1999**.
- [46] F. Fernández, X. García-Mera, M. Morales, J. E. Rodríguez-Borges, *Synthesis* **2001**, *2*, 239.
- [47] M. I. Nieto, J. M. Blanco, O. Caamaño, F. Fernández, X. García-Mera, J. Balzarini, E. Padalko, J. Neytsb, E. De Clercq, *Nucleosides Nucleotides* **1998**, *17*, 1255.
- [48] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693.
- [49] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. M. van de Sandt, W. Tong, G. Veith, C. Yang, *ATLA* **2005**, *33*, 155.

Received: September 10, 2009

Accepted: January 25, 2010

Published online: March 5, 2010