Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 580–585
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

ANN-QSAR model for selection of anticancer leads from structurally heterogeneous series of compounds

Humberto González-Díaz [a,*], Isis Bonet [b,c], Carmen Terán [d], Erik De Clercq [e], Rafael Bello [b], Maria M. García [b], Lourdes Santana [a], Eugenio Uriarte [a]

[a] *Department of Organic Chemistry, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain*
[b] *CEI, UCLV, Santa Clara 54830, Cuba*
[c] *Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Belgium*
[d] *Department of Organic Chemistry, University of Vigo, 36200 Vigo, Spain*
[e] *Katholieke Universiteit Leuven, Rega Institute, Minderbroedersstraat 10, B-3000 Leuven, Belgium*

Received 16 September 2006; received in revised form 29 November 2006; accepted 30 November 2006
Available online 15 December 2006

Abstract

Developing a model for predicting anticancer activity of any classes of organic compounds based on molecular structure is very important goal for medicinal chemist. Different molecular descriptors can be used to solve this problem. Stochastic molecular descriptors so-called the MARCH-INSIDE approach, shown to be very successful in drug design. Nevertheless, the structural diversity of compounds is so vast that we may need non-linear models such as artificial neural networks (ANN) instead of linear ones. SmartMLP-ANN analysis used to model the anticancer activity of organic compounds has shown high average accuracy of 93.79% (train performance) and predictability of 90.88% (validation performance) for the 8:3-MLP topology with different training and predicting series. This ANN model favourably compares with respect to a previous linear discriminant analysis (LDA) model [H. González-Díaz et al., J. Mol. Model 9 (2003) 395] that showed only 80.49% of accuracy and 79.34% of predictability. The present SmartMLP approach employed shorter training times of only 10 h while previous models give accuracies of 70–89% only after 25–46 h of training. In order to illustrate the practical use of the model in bioorganic medicinal chemistry, we report the *in silico* prediction, and *in vitro* evaluation of six new synthetic tegafur analogues having IC_{50} values in a broad range between 37.1 and 138 μg mL^{-1} for leukemia (L1210/0) and human T-lymphocyte (Molt4/C8, CEM/0) cells. Theoretical predictions coincide very well with experimental results.

Keywords: QSAR; Anticancer activity; ANN; Tegafur analogues; Linear discriminant analysis

1. Introduction

Quantitative structure–activity relationships (QSARs) have emerged as a rational alternative in order to find new active molecules including anticancer compounds [1,2]. Many topological molecular descriptors can be used to describe organic molecular structure with QSAR aims. Almost of them, for instance the Wiener index *W*, Harary number *H*, Randic invariant χ, and Balaban index *J* may be expressed as vector–matrix–vector forms. More recently, other interesting topologic indices such as the so-called Marrero-Ponce quadratic indices $q_k(X)$ have been introduced [3–7]. In particular, spectral moments of different matrices have been of special relevance for QSAR and other studies such as Gutman and Rosenfield studies on polymers' graphs, mean hydrophobicity moment, and the I3 index for proteins [8,9]. The success of the method of moments on the field of polymers has also been confirmed after Gónzalez M.P., *et al.* work on the analysis of the mutagenic power of dental polymers [10], which preceded the very related and recent work of Morales, *et al.* [11]. In general, the method of spectral moments has been

* Corresponding author.
E-mail addresses: gonzalezdiazh@yahoo.es, qohumbe@usc.es (H. González-Díaz).

0223-5234/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.11.016

largely used in many other different contexts of solid, theoretic, and bioorganic chemistry [12–14]. Gónzalez M.P. has additionally reported interesting applications of the method of moment in the molecular design of herbicides [15]. Other interesting applications of the method of moments in pharmaceutical sciences and medicinal chemistry were reported by Cabrera-Pérez *et al*. [16], and Molina *et al*., design of antibiotics [17]. Last but not least, several applications on medicinal chemistry of the method of moments developed by Estrada and Peña have appeared, including the design of sedative/hypnotic compounds and on the design of anticonvulsant drugs [18].

Particularly, our group has worked on a Markov model that uses stochastic spectral moments ${}^{SR}\pi_k$ to encode molecular structure with applications in nucleic acids, proteins and medicinal chemistry research. In any case, it has been demonstrated using vector–matrix–vector formalism that all these indices, including spectral moments too, are in fact very similar [19–25].

On the other hand, the search of anticancer compounds has always been on the desktop of molecular modelling and drug design specialists. In spite of this intensive search, the discovery of selective antitumor compounds has remained a largely elusive goal of cancer research. Subsequently, new approaches are needed in order to make an efficient search for candidates to be assayed as anticancer drugs [26–31]. Further, in the present work we have compared the LDA analysis previously reported by our group [26] with new one based on an ANN approach [32–34] using the database of 961 chemical compounds. In brief, first we have calculated 11 stochastic spectral moments ${}^{SR}\pi_k$ for each compound, followed by training and validation of new ANN using this data and finally testing the newly synthesized molecules selected using the derived model in order to prove the usefulness of model.

2. Results and discussion

2.1. Model train and validation

Table 1 depicts experimental results obtained from training a network by several types of MLP, with their time of training and the learning percentage. The results show clearly that SmartMLP version 1.5 increases learning rate while decreases required training time even if we use combined selection techniques or a learning coefficient. In order to speed up and optimize the ANN training process, SmartMLP version 1.5 enhances heuristics implemented in older versions. The learning coefficient was made a function on the size of the training set and the size of the samples in each class. Patterns selection heuristic over training set was improved gathering together some techniques implemented in older versions. The training algorithm was modified to use both uniform and RUL techniques in a way that it starts applying RUL until an error threshold was reached ($e < 0.1$) and then it applies uniform selection [32,33].

Table 1
Variants of MLP, time of training and learning percent

Backpropagation type	Training time (h)	Learning (%)
Traditional	46:50:00	70
SmartMLP version 1.0	30:00:00	82
Using learning coefficient	25:00:00	89
Combining selection techniques (RUL and Uniform)	25:00:00	87
Combining both previous (SmartMLP version 1.5)	10:00:00	94.04

SmartMLP helped selecting an adequate network topology from several candidates.

In order to optimize the performance of the method we trained the ANN using different topologies with one or two hidden layers and a variable number of neurons. SmartMLP 1.5 exhibited the best differentiation of anticancer from non-active compounds with 94.04% of accuracy. This topology presented two hidden layers having eight and three neurons, respectively, see Table 2. Changing network topology is a way to improve the QSAR results [34].

Finally, we have tested the predictability and robustness of the present ANN model by a re-substitution approach, which consists of different data partition in interchangeable training and predicting series [35]. SmartMLP 1.5 showed for all partitions an average training accuracy of above 93.79% and similar mean predictability of 90.88% in predicting series. Conversely, the LDA model previously reported suffers a lack of accuracy and predictability from 90.46% and 86.07% (partition 3) to averages of 80.49% and 79.34%, respectively, after re-substitution approach. These results have shown without doubt, the higher robustness of the present ANN models with respect to the LDA one [35], see Table 3. The names, ${}^{SR}\pi_k$ values, observed, and predicted classification of all the compounds used to train and test the model appears in supplementary material file, where the compounds misclassified are highlighted in red.

2.2. An example on the use of the model

The ANN model developed was used to predict the biological activity of some cyclopentylpirimidine derivative analogues of tegafur, a largely known anticancer drug (Fig. 1).

Table 2
SmartMLP classification percent for different topologies of one and two hidden layers

	Topologies[a]		SmartMLP 1.5 accuracy (%)
One hidden layer	2		70.98
	5		85.30
	6		87.58
	8		88.60
Two hidden layers	First layer	Second layer	
	5	2	90.50
	6	3	91.40
	8	2	93.12
	8	3	94.04

[a] Number of neurons in the hidden layers.

Table 3
Comparison of ANN and LDA classification results for several partitions of the dataset

	Sets	SmartMLP 1.5 learning (%)	LDA accuracy (%)
Partition 1	Training set	93.48	77.5
	Test set	91.67	78.3
Partition 2	Training set	93.62	77
	Test set	90.83	76.3
Partition 3	Training set	94.01	90.46
	Test set	89.75	86.07
Partition 4	Training set	94.04	77
	Test set	91.25	76.7
Mean	Training set	93.79	80.49
	Test set	90.88	79.34

We predicted the biological activity only for simple analogues from which CH_3, F, Cl, Br, and I substituted at the 5-carbon of the uracil skeleton was selected (Figs. 2 and 3). This selection was based on the simplicity and synthetic accessibility. All these analogues were synthesized in enough quantity for biological assay. Compounds **1**–**3** were prepared by condensation of the trimethylsilylated base (uracil, thymine, and fluorouracil) with bromocyclopentane in 11%, 10%, and 13% yield, respectively (Fig. 2) [36,37]. Compounds **4**–**6** were prepared from **1** using *N*-chlorosuccinimide or *N*-bromosuccinimide in acetic acid, or I_2 in nitric acid and dioxane, with 68%, 93%, and 94% yield, respectively (Fig. 3) [36].

The predictions coincide with the biological test result where all the compounds showed detectable biological activity in the three studied cellular lines namely L1210/0, Molt4/C8, and CEM/0 (Table 4). The most interesting activity on the leukemia line L1210/0 was presented by the F-tegafur analogue (compound **3**). However, results in the human lymphocyte lines were more discrete. Interestingly, the Br-tegafur analogue (compound **5**) resulted more selective for the human lymphocyte lines than the leukemia line. This result confirms the potentialities of the MARCH-INSIDE approach to model biological data and guide drug discovery in bioorganic medicinal chemistry [38,39].

3. Conclusions

We can conclude that the combination of MARCH-INSIDE with SmartMLP-ANN correctly classifies several anticancer compounds from heterogeneous series. This new strategy showed higher accuracy and robustness than LDA-based strategies. MARCH-INSIDE and SmartMLP-ANN approach was able to predict anticancer activity on leukemia and lymphocyte lines for pyrimidine derivatives, experimentally confirmed. All these conclusions coincide with multiple applications of Markov process in the literature [38–42].

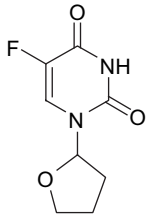

Fig. 1. Tegafur.

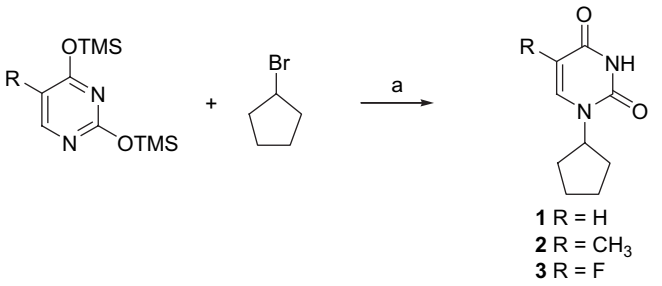

Fig. 2. Reagents and conditions: (a) i: CH_2Cl_2, reflux; ii: MeOH/H_2O (6:1).

4. Experimental section

4.1. The stochastic spectral moments

A precise definition of the descriptors generated by this methodology can be found in several reports of its application in the study of several biological properties. Briefly, the method uses stochastic or Markov matrix $^1\mathbf{\Pi}$ as a source of molecular descriptor. The Markov matrix is built up as a squared matrix $n \times n$ (n number of atoms in the molecule) whose elements ($^1p_{ij}$) are calculated as the ratio between the Pauling's electronegativity (χ_j) of the jth atom and the sum of the χ_k values for all the atoms covalently linked to the ith atom, including itself [40].

$$^1p_{ij} = \frac{\chi_j \bullet e^{\omega}}{\sum_{k=1}^{\delta+1} \chi_k \bullet e^{\omega}} \quad (1)$$

Local 3D characteristics of each atom are codified throughout the dummy variable ω_j. This variable (ω_j) takes the value $\omega_j = 1$ if the atom a_j is R, E or axial and takes the values $\omega_j = 0$, -1 whether the atom has no specific 3D properties or is S, Z or equatorial. The symbols R, S refer to the chirality of the atom. Alternatively, Z–E regards to the 3D

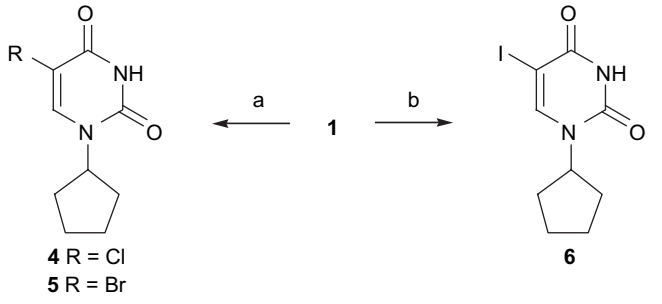

Fig. 3. Reagents and conditions: (a) NXS (X = Cl, Br), AcOH, reflux; (b) I_2, HNO_3, dioxane, 100 °C.

Table 4
Predicted, observed class and IC_{50} values for tegafur derivatives

Compound	Predicted class	Observed class	IC_{50} ($\mu g\,mL^{-1}$)		
			L1210/0	Molt4/C8	CEM/0
1	1	1	66.3	50.3	98.8
2	1	1	85.2	98.8	78.1
3	1	1	35.1	108	81.3
4	1	1	103	138	97.2
5	1	1	66.0	37.1	53.2
6	1	1	82.0	60.9	80.0

characteristic for atoms involved in double bonds [41]. In classical Markov theory, these numbers are the probabilities with which the system returns to the initial state. In the present context, they are the probabilities with which electrons return to the atoms at different distances after an arbitrary initial observation time t_0 [42]. The calculation of ${}^{SR}\pi_k$ for any organic or inorganic molecule was carried out using the MARCH-INSIDE software [43].

$$ {}^{SR}\pi_k = \sum_{i=1}^{n} {}^{k}p_{ii} \tag{2} $$

4.2. Heuristics added to the platform SmartMLP for ANN training

Artificial neural networks have been used to solve numerous types of problems. Classification is one of the kinds of problems where they are commonly applied. The backpropagation (BP) algorithm is one of the most popularly applied to feed-forward training of neural networks, due to its simplicity, its capacity to extract useful information from samples and to store it implicitly as weights over their connections. However, this algorithm also has its limitations of practical order, which are generally accepted and studied by researchers. Some of these limitations are (1) its convergence towards a state of minimum error can be extremely slow, principally if the size of the network is not very big in oppose to the problem to manipulate, (2) it can stagnate in local minima before finishing the learning of all samples, (3) it is almost impossible to select the network design a priori [32–34].

Improving the BP algorithm constitutes currently the work of researchers worldwide. SmartMLP version 1.5 introduces some improvements to BP training algorithm to enhance its performance [44]. Learning coefficient (η) as the following one was able to accelerate the training process.

$$ \eta = \frac{C}{2} \times \frac{N_1 + N_2 + \cdots + N_m}{N_1^2 + N_2^2 + \cdots + N_m^2} \tag{3} $$

where C is the number of examples of the training set, m the number of classes and N_i is the number of examples that lie within the ith class. The frequency or order of the patterns presentation also affects the training of the network. SmartMLP has implemented three types of selection patterns:

1 Uniform: In this selection type, each pattern is selected randomly but the probability of being selected is the same one for each pattern.
2 Sequentially: The examples here are selected in the same order that appear in the training set.
3 Repeat until learning (RUL): This is a pedagogic selection type strategy. Each example is presented to the network depending on the error the network commits classifying it. So that each example is selected randomly and repeated until its error is below the error average made by the network increased by a certain factor β.

$$ \beta = d\frac{1}{M}\sum_{u=1}^{M} (\varepsilon^u)^2 \tag{4} $$

where $d = 1.5$ and M is the number of examples, that implies repetition of only examples with error values substantially higher than the mean. It is not very easy to find a good initial weights set that facilitates the network training. The union of two of these strategies (RUL and Uniform) of selection of parameters gave very good results [44].

4.3. Design and implementation of the neural network

We have a knowledge base of 961 examples, 298 structurally heterogeneous anticancer compounds, and 664 non-anticancer compounds. Active compounds (labelled as 1) are those reported as anticancer compounds by Martin Negwer [45]or Kleman *et al.* [46]. By the contrary, non-anticancer compounds (labelled as 0) are selected at random from the same database. Several neural networks were implemented with one and two hidden layers. We use a combination of strategies for selecting parameters: RUL and Uniform, as well as the learning coefficient previously mentioned. Finally the dataset was divided, choosing a control sample in a random way, equivalent to 25% of the total sample. This was repeated several times and with these datasets different neural networks were built, testing with different network topologies. Finally, we obtain a neural network with a 94.04% of accuracy for the training set and 91.25% for the external control set [35].

4.4. Chemistry

Melting points were determined using a Reichert Kofler thermopan or in capillary tubes on a Büchi 510 apparatus and are uncorrected. IR spectra were recorded on a Perkin–Elmer 1640FT spectrophotometer. ^1H and ^{13}C NMR spectra were recorded on a Bruker AMX spectrometer at 300 and 75.47 MHz, respectively, using TMS as internal standard (chemical shifts in δ values, J in hertz). Mass spectra were obtained using a Hewlett–Packard 5988A spectrometer. Elemental analyses were performed using a Perkin Elmer 240B microanalyzer and were within −0.4 to +0.4% of calculated values in all cases. 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 F_{254}, 0.25 mm).

4.4.1. Condensation of pyrimidine bases with cyclopentyl bromide. General procedure

A mixture of pyrimidine (1.53 mmol) and a catalytic amount of ammonium sulphate in hexamethyldisilazane (12 mL) was heated at 135 °C for 12 h under an Ar atmosphere. The resulting clear solution was concentrated *in vacuo* under anhydrous conditions to yield the silylated pyrimidine as a colourless oil. This oil was immediately dissolved in dry 1,2-dichloroethane (4 mL), and a solution of cyclopentyl bromide (1.53 mmol) in the same solvent (4 mL) was added. This mixture was heated to reflux under an inert atmosphere (48 h for compounds **1** and **2**, and one week for compound **3**) and then cooled, treated with 6:1 methanol/water and filtered. The filtrate was extracted with CH_2Cl_2 and the solution obtained was dried over anhydrous Na_2SO_4. The drying agent was filtered out and the solvent was evaporated *in vacuo*; the residue was purified by FC using 3:1 hexane/ethyl acetate as an eluent.

*4.4.1.1. 1-Cyclopentyluracil (**1**).* Compound **1** was prepared from uracil in 11% yield, m.p. 170–172 °C. IR: 3000, 2875, 2826, 1678, 1613, 1470, 1421, 1383, 1267. ^{1}H NMR ($CDCl_3$): 8.13 (1H, s, NH), 7.22 (1H, d, H-6, $J = 8.1$), 5.72 (1H, dd, H-5, $J = 8.1$ and 2.3), 4.90 (1H, m, H-1′), 2.12 (2H, m, 1H-2′ + 1H-5′), 1.90–1.50 (6H, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 163.4, 151.3, 141.3, 102.8, 57.1, 31.8, 24.5. MS *m/z* (%): 180 (M^+, 19), 113 ($[Ura + 1]^+$, 100), 112 (Ura^+, 16), 69 (11). Anal. $C_9H_{12}N_2O_2$ (C, H, N).

*4.4.1.2. 1-Cyclopentyl-5-methyluracil (**2**).* Compound **2** was prepared from 5-methyluracil in 10% yield, m.p. 174–176 °C. IR: 3174, 3038, 2956, 1690, 1659, 1474, 1269. ^{1}H NMR ($CDCl_3$): 9.10 (1H, s, NH), 7.01 (1H, q, H-6, $J = 1.1$), 4.91 (1H, m, H-1′), 2.10 (2H, m, 1H-2′ + 1H-5′), 1.93 (3H, d, CH_3, $J = 1.1$), 1.83–1.55 (6H, m, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 164.1, 151.5, 137.1, 111.3, 56.7, 31.6, 24.5. MS *m/z* (%): 195 ($[M + 1]^+$, 6), 127 ($[Thy + 1]^+$, 63), 126 (Thy^+, 100), 83 (16). Anal. $C_{10}H_{14}N_2O_2$ (C, H, N).

*4.4.1.3. 1-Cyclopentyl-5-fluorouracil (**3**).* Compound **3** was prepared from 5-fluorouracil in 13% yield, m.p. 178–180 °C. IR: 3169, 3048, 1710, 1670, 1659, 1261. ^{1}H NMR ($CDCl_3$): 9.45 (1H, s, NH), 7.27 (1H, d, H-6, $J = 6.3$), 4.93 (1H, m, H-1′), 2.10 (2H, m, 1H-2′ + 1H-5′), 1.85–1.50 (6H, m, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 157.4 and 157.1 (d, C_4, $J = 27$), 150.3 (C_2), 142.6 and 139.5 (d, C_5, $J = 237$), 125.8 and 125.3 (d, C_6, $J = 32$), 57.4, 31.7, 24.4. MS *m/z* (%): 199 ($[M + 1]^+$, 4), 198 (M^+, 25), 131 ($[FUra + 1]^+$, 100), 130 ($FUra^+$, 30). Anal. $C_9H_{11}FN_2O_2$ (C, H, N).

*4.4.2. Halogenation of **1** using N-halosuccinimide. General procedure*

To a solution of **1** (100 mg, 0.55 mmol) in acetic acid (3 mL) was added a solution of *N*-halosuccinimide (0.60 mmol) in acetic acid (6 mL), and the mixture was heated to reflux (48 h for compound **4**, and 6 h for compound **5**). After the reaction mixture had cooled, the solvent was evaporated and the residue was purified by FC using 4:1 hexane/ethyl acetate as an eluent.

*4.4.2.1. 5-Chloro-1-cyclopentyluracil (**4**).* Prepared from **1** and *N*-chlorosuccinimide in 68% yield, m.p. 196–198 °C. IR: 3167, 3031, 2967, 1717, 1661, 1613, 1471, 1316. ^{1}H NMR ($CDCl_3$): 8.90 (1H, s, NH), 7.45 (1H, s, H-6), 4.90 (1H, m, H-1′), 2.15 (2H, m, 1H-2′ + 1H-5′), 1.95–1.55 (6H, m, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 159.4, 150.8, 138.5, 109.3, 57.8, 31.9, 24.4. MS *m/z* (%): 216 ($[M + 2]^+$, 10), 214 (M^+, 30), 149 ($[ClUra + 2]^+$, 35), 148 ($[ClUra + 1]^+$, 30), 147 ($[ClUra]^+$, 100), 146 ($[ClUra - 1]^+$, 72), 103 (25), 69 ($[M - ClUra]^+$, 12), 68 ($[M - ClUra - 1]^+$, 10), 67 ($[M - ClUra - 2]^+$, 15). Anal. $C_9H_{11}ClN_2O_2$ (C, H, N).

*4.4.2.2. 5-Bromo-1-cyclopentyluracil (**5**).* Prepared from **1** and *N*-bromosuccinimide in 93% yield, m.p. 200–202 °C. IR: 3160, 3028, 2965, 2840, 1718, 1654, 1606, 1466, 1272. ^{1}H NMR ($CDCl_3$): 8.82 (1H, s, NH), 7.52 (1H, s, H-6), 4.90 (1H, m, H-1′), 2.14 (2H, m, 1H-2′ + 1H-5′), 1.90–1.50 (6H, m, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 159.6, 151.1, 141.1, 97.1, 57.9, 31.9, 24.4. MS *m/z* (%): 260 ($[M + 2]^+$, 36), 258 (M^+, 36), 193 ($[BrUra + 3]^+$, 100), 192 ($[BrUra + 2]^+$, 91), 191 ($[BrUra + 1]^+$, 98), 190 ($BrUra^+$, 84), 149 (33), 147 (30), 69 ($[M - BrUra]^+$, 13), 68 ($[M - BrUra - 1]^+$, 11), 67 ($[M - BrUra - 2]^+$, 16), 53 (13). Anal. $C_9H_{11}BrN_2O_2$ (C, H, N).

*4.4.2.3. 1-Cyclopentyl-5-iodouracil (**6**).* A mixture of **1** (93 mg, 0.52 mmol), iodine (260 mg, 1.04 mmol) and 0.75 M nitric acid (0.68 mL) in dioxane (8 mL) was stirred for 2 h at 100 °C. After the reaction mixture had cooled, the solvent was evaporated *in vacuo* and the residue was purified by FC using 4:1 hexane/ethyl acetate as an eluent which gave **6** (148 mg, 94% yield), m.p. 220–222 °C. IR: 3148, 1955, 1839, 1699, 1664, 1599, 1427, 1271. ^{1}H NMR ($CDCl_3$): 9.09 (1H, s, NH), 7.61 (1H, s, H-6), 4.88 (1H, m, H-1′), 2.13 (2H, m, 1H-2′ + 1H-5′), 1.90–1.53 (6H, m, 1H-2′ + H-3′ + H-4′ + 1H-5′). ^{13}C NMR ($CDCl_3$): 160.3, 151.1, 146.2, 68.3, 57.9, 31.9, 24.4. MS *m/z* (%): 306 (M^+, 43), 239 ($[IUra + 1]^+$, 48), 238 ($IUra^+$, 100), 195 (34). Anal. $C_9H_{11}IN_2O_2$ (C, H, N).

4.5. Biological activity data and assay

All assays were performed in a flat-bottomed 96-well microtiter plates. To each well were added 5×10^4 L1210/0, or 7.5×10^4 CEM/0, or Molt4/C8 cells and a certain amount of the test compound. The cells were allowed to proliferate for 48 h (L1210/0) or 72 h (CEM/0, 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 the compound that reduced the number of living cells by 50% [47].

Acknowledgements

The authors thank funds from Xunta de Galicia projects PXIB20304PR and BTF20302PR as well as from Ministerio de Sanidad y Consumo project PI061457. The authors are also indebted to Prof. Dr. Rubio-González A. (Post-graduation and Research, UCLV) and Prof. Dr. Morales S. (VLIR project coordinator Faculty of Pharmacy, UCLV) by kind attention and partial assignation of Vlaamse Interuniversitaire Raad (VLIR) USO, University Development Cooperation program by partial financial support. González-Díaz, H. acknowledges contract as guest-professor from the Department of Organic Chemistry of the Faculty of Pharmacy at the University of Santiago de Compostela, Spain in 2006. Bonet, I. acknowledges hospitality and financial support from Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Belgium.

Appendix. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.ejmech.2006.11.016.

References

[1] H. Kubinyi, J. Taylor, C. Ramdsen, Quantitative Drug Design, in: C. Hansch (Ed.), Comprehensive Medicinal Chemistry, vol. 4, Pergamon, 1990, pp. 589–643.
[2] E. Estrada, E. Uriarte, Curr. Med. Chem. 8 (2001) 1573.
[3] H. Wiener, J. Am. Chem. Soc. 69 (1947) 17.
[4] J. Devillers, A.T. Balaban, Topological Indices and Related Descriptors in QSAR and Drug Design, Gordon & Breach, Amsterdam, 2000, pp. 3–41.
[5] L.B. Kier, L.H. Hall, Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach Sci. Pub., Amsterdam, 1999, pp. 455–489.
[6] Y. Marrero-Ponce, H. González-Díaz, V. Romero-Zaldivar, F. Torrens, E.A. Castro, Bioorg. Med. Chem. 12 (2004) 5331.
[7] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, Germany, 2000.
[8] I. Gutman, V.R. Rosenfield, Theor. Chim. Acta 93 (1996) 191.
[9] E. Estrada, Bioinformatics 18 (2002) 1.
[10] M.P. González, A.H. Morales, R. Molina, Polymer 45 (2004) 2773.
[11] A.H. Morales, M.P. González, J.B. Rieumont, Polymer 45 (2004) 2045.
[12] J.K. Burdett, S. Lee, J. Am. Chem. Soc. 107 (1985) 3063.
[13] S. Markovic, I. Gutman, J. Mol. Struct. (Theochem.) 81 (1991) 81.
[14] E. Estrada, H. González-Díaz, J. Chem. Inf. Comput. Sci. 43 (2003) 75.
[15] M.P. González, H. González-Díaz, R. Molina, M.A. Cabrera-Pérez, A.R. Ramos de, J. Chem. Inf. Comput. Sci. 43 (2003) 1192.
[16] M.A. Cabrera-Pérez, H. González-Díaz, T.C. Fernandez, J.M. Pla-Delfina, S.M. Bermejo, Eur. J. Pharm. Biopharm. 53 (2002) 317.
[17] E. Molina, H. González-Díaz, M.P. González, E. Rodríguez, E. Uriarte, J. Chem. Inf. Comput. Sci. 44 (2004) 515.
[18] E. Estrada, A. Peña, Bioorg. Med. Chem. 8 (2000) 2755.
[19] H. González-Díaz, R. Uriarte, A.R. Ramos de, Bioorg. Med. Chem. 13 (2004) 323.
[20] H. González-Díaz, E. Olazábal, N. Castañedo, S.I. Hernádez, A. Morales, H.S. Serrano, J. González, A.R. Ramos de, J. Mol. Model. 8 (2002) 237.
[21] H. González-Díaz, O. Gia, E. Uriarte, I. Hernádez, R. Ramos, M. Chaviano, S. Seijo, J.A. Castillo, L. Morales, L. Santana, D. Akpaloo, E. Molina, M. Cruz, L.A. Torres, M.A. Cabrera, J. Mol. Model. 9 (2003) 395.
[22] H. González-Díaz, S.I. Hernández, E. Uriarte, L. Santana, Comput. Biol. Chem. 27 (2003) 217.
[23] H. González-Díaz, A.R. Ramos de, R.R. Molina, Bull. Math. Biol. 65 (2003) 991.
[24] H. González-Díaz, A.R. Ramos de, E. Uriarte, Online J. Bioinf. 1 (2002) 83.
[25] A.R. Ramos de, H. González-Díaz, R. Molina, M.P. González, E. Uriarte, Bioorg. Med. Chem. 12 (2004) 4815.
[26] E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana, E. De Clercq, J. Med. Chem. 43 (2000) 1975.
[27] W.A. Denny, The role of medicinal chemistry in the discovery of DNA-active anticancer drugs: from random searching, through lead development, to the novo design, in: M.J. Waring, B.A.J. Ponder (Eds.), The Search for Anticancer Drugs, Kluwer, Dordrecht, 1992 (chapter 2).
[28] H. Kubinyi, J. Recept. Signal Transduct. Res. 19 (1999) 15.
[29] W.P. Walters, M.T. Stahl, M.A. Murcko, Drug Discov. Today 3 (1998) 160.
[30] K. Ferrante, B. Winograd, R. Canetta, Cancer Chemother. Pharmacol. 43 (1999) S61.
[31] E. Menta, M. Palumbo, Expert Opin. Ther. Pat. 8 (1998) 1627.
[32] I. Bonet-Cruz, Learning Optimization in a MLP Neural Network Applied to OCR. MICAI 2002: Advances in Artificial Intelligence, LNAI, vol. 2313, 2002, pp. 292–300.
[33] C. Caching, Neural Netw. 7 (1994) 175.
[34] S. Vilar, L. Santana, E. Uriarte, J. Med. Chem. 49 (2006) 1118.
[35] H. Van Waterbeemd, Discriminant analysis for activity prediction, in: R. Manhnhold, Krogsgaard-Larsen, H. Timmerman, H. Van Waterbeemd (Eds.), Method and Principles in Medicinal Chemistry, Chemometric Methods in Molecular Design, vol. 2, VCH, Weinhiem, 1995, pp. 265–282.
[36] C. Teran, M. Teijeira, L. Santana, E. Uriarte, A. Castiñeiras, J. Mol. Struct. 448 (1998) 69.
[37] C. Teran, L. Santana, M. Teijeira, E. Uriarte, J. Balzarini, E. De Clercq, Pharmazie 53 (1998) 644.
[38] H. González-Díaz, R.R. Molina, E. Uriarte, Bioorg. Med. Chem. Lett. 14 (2004) 4691.
[39] H. González-Díaz, A.R. Ramos de, R.R. Molina, Bioinformatics 19 (2003) 2079.
[40] L. Pauling, The Nature of Chemical Bond, Cornell University Press, Ithaca, New York, 1939, pp. 2–60.
[41] E. Eliel, S. Wilen, L. Mander, Stereo Chemistry of Organic Compounds, John Wiley & Sons Inc., New York, 1994, pp. 103–112.
[42] H. González-Díaz, I. Bastida, N. Castañedo, O. Nasco, E. Olazabal, A. Morales, H.S. Serrano, A.R. Ramos de, Bull. Math. Biol. 66 (2004) 1285.
[43] H. González-Díaz, I. Hernández, MARCH-INSIDE 2002, version 1.0 (MARkovian CHemicals IN SIlico DEsign). This is a preliminary experimental version. Future professional version shall be available to the public. For any information about it send an e-mail to the corresponding author. <mailto:gonzalezdiazh@yahoo.es>.
[44] I. Bonet-Cruz, R. Bello, M.M. García, SmartMLP version 1.5, 2003.
[45] M. Negwer, Organic Chemical Drugs and Their Synonyms, Akademie-Verlag, Berlin, 1987.
[46] A. Kleeman, J. Engel, B. Kutscher, D. Reichert, Pharmaceutical Substances, fourth ed. George Thieme Verlag, Stuttgart, 2001.
[47] E. De Clercq, in: In vivo and Ex vivo Test Systems to Rationalize Drug Design and Delivery, Editions Sante, Paris, France, 1994, pp. 108–125.