Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 4999–5006

BIOORGANIC & MEDICINAL CHEMISTRY

# Synthesis and QSAR Study of the Anticancer Activity of Some Novel Indane Carbocyclic Nucleosides

S.-W. Yao,[a] V. H. C. Lopes,[a] F. Fernández,[b] X. García-Mera,[b,*] M. Morales,[b] J. E. Rodríguez-Borges[c] and M. N. D. S. Cordeiro[a,*]

[a]*REQUIMTE/Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

[b]*Departamento de Química Orgánica, Facultade de Farmacia, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

[c]*CIQ/Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

Received 7 July 2003; accepted 1 September 2003

**Abstract**—A set of 14 indane carbocyclic nucleosides were synthesized and experimentally assayed for their inhibitory effects in the proliferation of murine leukemia (L1210/0) and human T-lymphocyte (Molt4/C8, CEM/0) cells. The compounds have promising inhibitory activity judging from the $IC_{50}$ values obtained for all these cellular lines. Multiple linear regression analysis was then applied to build up consistent QSAR models based on quantum mechanics-derived molecular descriptors. The derived models reproduce well the experimental data of both three cells ($r^2 \geq 0.90$), display a good predictive power ($r^2_{CV} \geq 0.86$) and are, above all, easily interpretable. They show that frontier-orbital energies and hydrophobicity are mainly responsible for the activity of the synthesized compounds and also, suggest similar mechanisms of action. The final QSAR-models involve only two descriptors: the lowest unoccupied molecular orbital energy and the solvent accessible-hydrophobic surface area, but describe a sound correlation between predicted and experimental activity data ($r^2 = 0.931$, $r^2 = 0.936$ and $r^2 = 0.931$ for the cells L1210/0, Molt4/C8 and CEM/0, respectively).


## Introduction

Carbocyclic analogues of nucleosides (CANs) are compounds in which the furan ring of the nucleoside is replaced by a carbocyclic system. This modification makes carbocyclic nucleosides more resistant to hydrolases than the natural nucleosides themselves, but does not prevent their enzymatic conversion to nucleotide analogues. CANs are thought to inhibit the DNA synthesis, that is they act as chain terminators in that, following their incorporation into the growing DNA chain, they do not allow further chain elongation.[1,2] As a result, they can possess antiviral, antineoplastic activity or other potentially biological properties.[3,4]

In the past, it has been thought important to improve the biological activity of CANs by decreasing their susceptibility to degradation by nucleoside phosphorylase.[5] This can be achieved, for instance, by a proper chemical modification of the sugar ring which does not deteriorate their activity. Following this line of thought, we have recently designed a number of derivatives of the carbocyclic nucleosides Carbovir and Abacavir.[6–8] The strong anticancer properties of these compounds prompted us to examine the structural/activity features of their aminoalcohol moiety and further, to modify it in an attempt to improve their liposolubility. To this end, 14 novel derivatives in which the cyclopentene ring is replaced by an indane system have been synthesized in one of our laboratories. Based on the hypothesis that selective in vitro activity of a compound against cancer cell lines can predict its activity against the corresponding human's tumors, the synthesized compounds have been assayed on different cellular lines, namely on murine leukemia (L1210/0) and human T-lymphocyte (Molt4/C8 and CEM/0) cells.

The interaction mechanisms of anticancer drugs are still generally not well understood at the molecular level. Actually, this is a very complex problem that certainly benefits from the alternative, more versatile view computational techniques, like Quantitative Structure–Activity Relationship (QSAR), are able to provide. In

*Corresponding authors. Fax: +34-981-594912; e-mail: qoxgmera@usc.es (X. García-Mera); fax: +351-22-6082959; e-mail: ncordeir@fc.up.pt(M. N. D. S. Cordeiro).

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.09.005

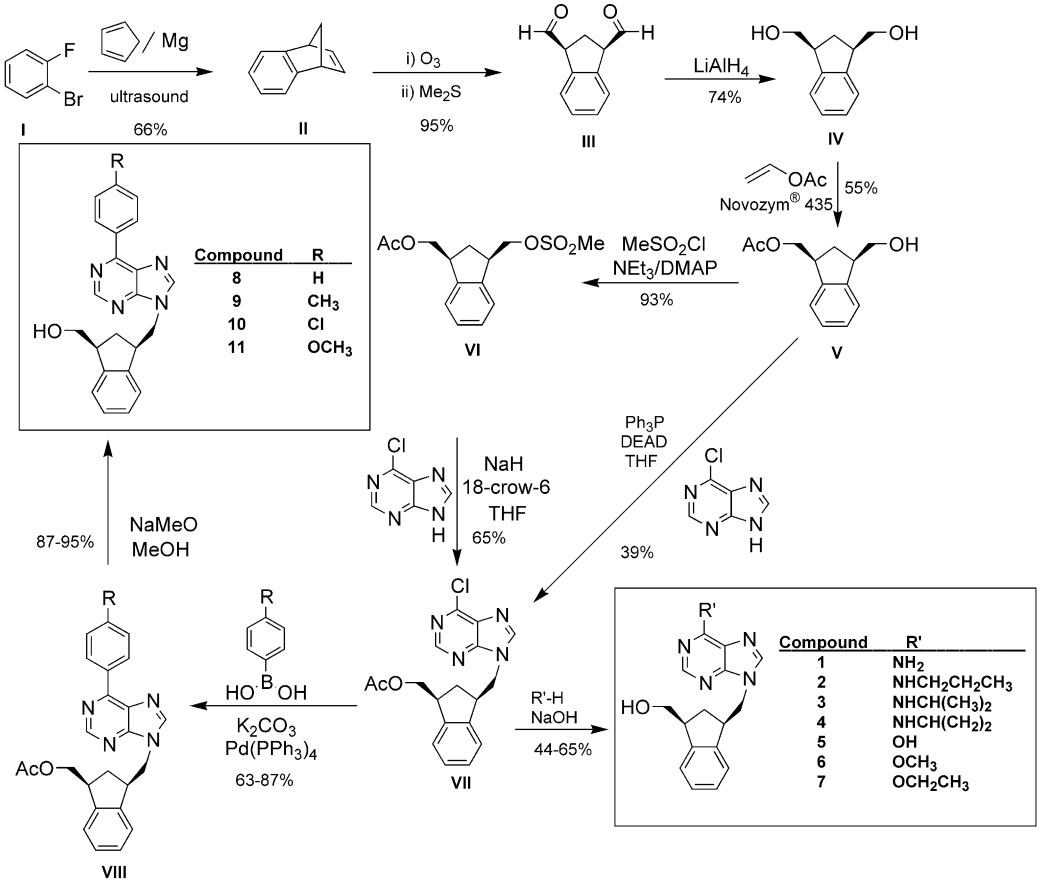
fact, a reliable predictive QSAR model may offer new insights into the drugs' action mode and additionally, guide molecular screening and design.[9–11]

The major assumption of any QSAR research is that structurally similar molecules have similar physicochemical properties and thus are thought to have similar biological action. Generally speaking, molecular properties like electronic, hydrophobic, steric, hydrogen bonding and dispersion properties may all be relevant to characterize the different working processes of drug action.[12,13] In some cases, all molecular descriptors are important but in other cases only few are determinant, and the remainders are playing only a minor part.[14] This is indeed the case in the present study, since we have found that electronic properties, in particular frontier-orbital energies, and hydrophobicity can correlate drug activity remarkably well.

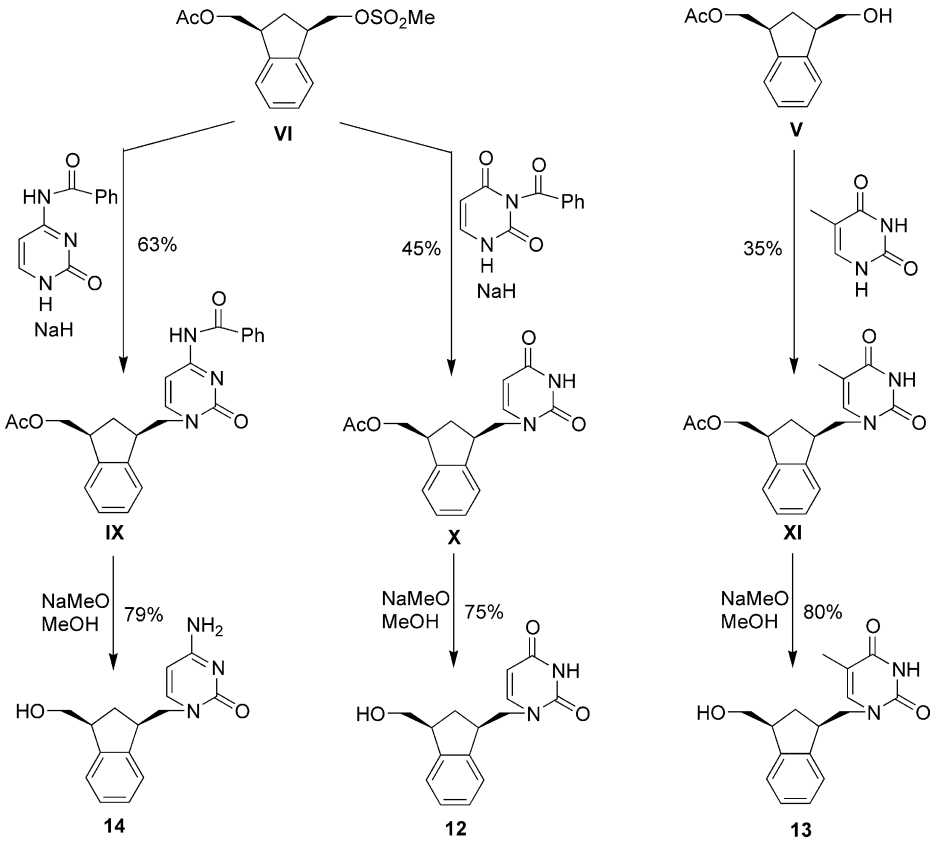
## Synthesis and Evaluation of Anticancer Activity

The preparation of compounds **1–11** (Scheme 1)[15] proceeds via generation of benzonorbornadiene (**II**) by means of the [4 + 2] cycloaddition of benzyne, generated in situ from 1-bromo-2-fluorobenzene, to cyclopentadiene in THF. Treatment of benzonorbornadiene with ozone and quenching with dimethyl sulphide afforded dialdehyde **III**, which upon treatment, without prior isolation, with lithium aluminium hydride gave diol **IV**. Two methods were then compared for protection of the left-hand hydroxyl of **IV**: reaction with acetic anhydride and pyridine, and transesterification with vinyl acetate catalysed by Novozym® 435. Both gave mixtures of **IV**, **V** and the diacetylated product with similar yields of **V** (around 55%), but with the enzymatic process workup was easier and the recovery of **IV** was greater. 6-Chloropurine was first condensed with **V** by a standard Mitsunobu reaction[16] in the presence of triphenylphosphine and diethylazodicarboxylate, but with this method we were unable to achieve yields greater than 40%. More satisfactory was an indirect route via mesylate *VI* (obtained from *V* by treatment at 0 °C with methanesulfonyl chloride, triethylamine and DMAP): reaction of crude *VI* with 6-chloropurine in DMF in the presence of NaH and 18-crown-6 ether gave the condensation product **VII**. Purinylcarbonucleosides **1–7**, in which the purine 6-substituent is an NHR′ or OR′ group, were obtained directly from **VII** by treatment with the corresponding nucleophiles $NH_2R'$ or R′OH. Treatment of **VII** with boronic acids under Suzuki conditions[17] using potassium carbonate as base, tetrakis(triphenylphosphine) palladium as catalyst and toluene as solvent and deprotection with NaMeO/MeOH gave compounds **8–11**.

Pyrimidine derivative **12** (Scheme 2)[18] was prepared by reaction of **VI** with 3-benzoyluracil in DMF in the presence of NaH and 18-crown-6 ether and deprotection with NaMeO/MeOH. By a similar strategy the cytosine analogue **14** was prepared from **VI** by reaction with *N*-




**Scheme 1.**




**Scheme 2.**

benzoylcytosine and subsequent deprotection. On the other hand, the attempted reaction of 3-benzoylthymine with **VI** was unsuccessful. More satisfactory was a standard Mitsunobu reaction. Treatment of **V** with thymine in the presence of triphenylphosphine and DEAD and similar deprotection, as described above, afforded thymine derivative **13**.

The synthesized indane derivatives were evaluated for their anti-tumor effects on the proliferation of murine leukemia cells (L1210/0) and human T-lymphocyte cells (Molt4/C8 and CEM/0). The inhibitory activity was estimated as $IC_{50}$, that is the concentration required to reduce cells' proliferation by 50%, according to the method of De Clercq.[9] Then, as usual, such data was changed to $-\log(IC_{50})$ to be of practical use in the following QSAR modeling. Table 1 presents the experimental data for the inhibitory activity of the synthesized compounds.

## Procedure

### Computational methods

The structures of all compounds were first drawn with the aid of QUANTA software package,[19] running on a

**Table 1.** Indane derivatives' inhibitory activity[a] on three different cellular lines: L1210/0, Molt4/C8 and CEM/0

| Compd | L1210/0 | | Molt4/C8 | | CEM/0 | |
|---|---|---|---|---|---|---|
| | $IC_{50}\pm SEM$ | $-\log(IC_{50})$ | $IC_{50}\pm SEM$ | $-\log(IC_{50})$ | $IC_{50}\pm SEM$ | $-\log(IC_{50})$ |
| **1** | 100±5 | −2.53 | 81±1 | −2.44 | 82±2 | −2.44 |
| **2** | 71±1 | −2.32 | 37±2 | −2.04 | 49±3 | −2.16 |
| **3** | 54±9 | −2.20 | 24±1 | −1.85 | 29±2 | −1.93 |
| **4** | 85±7 | −2.40 | 58±2 | −2.24 | 68±0 | −2.31 |
| **5** | 77±10 | −2.42 | 43±2 | −2.16 | 48±1 | −2.21 |
| **6** | 108±11 | −2.54 | 66±5 | −2.33 | 73±2 | −2.37 |
| **7** | 100±5 | −2.33 | 100±5 | −1.79 | 100±5 | −1.79 |
| **8** | 10±1 | −1.45 | 3.4±0.3 | −0.980 | 3.2±0.3 | −0.954 |
| **9** | 5.7±0.7 | −1.19 | 1.6±0.04 | −0.636 | 1.4±0.5 | −0.578 |
| **10** | 3.3±0.2 | −0.927 | 0.81±0.16 | −0.317 | 0.63±0.15 | −0.208 |
| **11** | 6.8±1.6 | −1.25 | 3.2±0.8 | −0.919 | 2.7±0.8 | −0.845 |
| **12** | >200 | −2.87 | >200 | −2.87 | >200 | −2.87 |
| **13** | 112±18 | −2.59 | 81±2 | −2.45 | 81±12 | −2.45 |
| **14** | >200 | −2.87 | >200 | −2.87 | >200 | −2.87 |

[a]Inhibitory activity estimated as $IC_{50}$ (50% growth-inhibitory concentration in μg/mL) and then transformed to $-\log(IC_{50})$.

Silicon Graphics $O^2$/R10000 workstation. Molecular structures were then fully optimized with the PM3 semi-empirical Hamiltonian[20] implemented in the GAUSSIAN98 program,[21] running on a Pentium IV, and electronic descriptors were extracted from the Gaussian files using home-developed scripts. Here, it should be remarked that the molecular structures pertain only to the compounds' global minimum energy conformation, and indeed, calculated averages from a Boltzmann distribution of the most populated conformations in solution would be desirable. But the point of any QSAR model is to have a set of readily calculated descriptors, and such an approach would require much more extensive calculations.

Subsequently, the optimized structures were retrieved back to QUANTA for further analysis. Therein, solvent-accessible Connolly surface areas[22,23] (i.e., the surface traced out by the surface of a probe water sphere rolling over the van der Waals molecular surface) and the corresponding enclosed molecular volumes were calculated for each compound. Final equation coefficients and statistical parameters were determined by multilinear stepwise regression with SPSS[24] using usual techniques for retaining variables, including cross-correlation and multicollinearity diagnostics.[25] In keeping with statistical standards, correlation equations were only considered to be acceptable if the (*adjusted*) product correlation coefficient, $r$, accounted for more than 80% of the variance ($r^2 \geq 0.80$).

## QSAR modelling

A QSAR modeling approach seeks to uncover correlation of biological activity with molecular structure. The relationship is most often expressed by a multilinear equation that relates molecular structural features to the desired activity:

$$\text{activity} = \sum_i x_i \alpha_i + b$$

where $x$ stands for the particular molecular feature (*descriptor*) and $\alpha$, $b$ are the parameters to be optimized using a known data set. As the name implies, a linear equation only identifies linear relationships between the molecular descriptors and the bioactivity to be predicted. However, in some cases extremely non-linear relationships exist and thus other more sophisticated methods have to be applied, such as artificial neural networks or genetic algorithms.[26]

Nowadays there is a vast amount of available molecular descriptors with which one can model the activity of interest. This complicates the task of selecting those that will be more suitable, especially when one tries to define an accurate, robust and (most importantly) interpretable model. Furthermore, as the compounds' interaction mechanisms are frequently unknown, the assumed QSAR models are to some extent arbitrary. Chance correlations are also to be expected coming from either an incorrect model formula or noise. We resorted here to a supervised variable selection modeling, carefully analyzing collinearity between variables and restraining the number of included variables to avoid possible spurious relationships.

When a chemical compound is administered to an organism, two events must occur for a biological response to be triggered. Firstly, the compound has to be transported to the site of action (*receptor*), and secondly, it must interact with the target in an adequate manner. To code these processes into a valid QSAR model, the following strategy has been pursued. To begin with, the molecular descriptors were roughly clustered into two groups. The first one—here referred to as **class I**—aims to describe the interaction effects. It includes quantum calculated electronic descriptors, such as the energies of the highest occupied molecular orbital (HOMO) and of the lowest unoccupied molecular orbital (LUMO), Mulliken atomic charges[27] or dipole moments. On the other hand, the second one—**class II**—tries to model the drug delivery and size effects, and contains the computed hydrophobic/hydrophilic solvent-accessible surface areas and molecular volumes, including 2D descriptors as the octanol–water partition coefficient. All these descriptors as well as certain combination of them (e.g., the energy gap between the HOMO and LUMO) were then taken as possible predictor variables in the multiple linear regressions for the outcome measure $-\log(IC_{50})$. Notice that such combinations might offset the linear approximation assumption of the model. Table 2 summarizes and defines the predictor variables used in this work.

## Results and Discussion

Following the strategy outlined before, we began by finding the most relevant descriptor variables for each class. Firstly, all descriptors that had a rather low correlation with activity ($|r| < 0.3$) were immediately discarded. Then, highly collinear descriptors ($|r| > 0.9$) were identified by examining the cross-correlation matrix. Descriptor $M_r$, for example, was found to be highly correlated to descriptors PHO ($r = 0.97$) and LogP ($r = 0.92$). Rather than deleting any of such descriptors, they were not included together in the regressions. Multilinear stepwise regressions were then carried out with the groups of variables considered. Even then, some suggested equations were rejected by mechanistic reasons. For instance, we refused a model that included only two descriptors from class II (e.g., for the L1210/0 and Molt4/C8 cells, an equation with only PHO and LogP), since it is quite improbable that electronic effects have no contribution to the activity.

The resulting best-fit models are given in Table 3 together with the statistical parameters of the regressions, while the values for the variables are in Table 4. Notice that these models were derived using the entire data set of compounds ($n = 4$) since no outliers were identified. The overall quality of the models is indicated by the correlation coefficient adjusted for degrees of freedom, $r^2$, the standard deviation, $s$, and the Fisher's statistic, $F$, including predictive power by the correlation coefficient

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

| Class | Symbol | Definition |
|---|---|---|
| **I** | LUMO | energy of the lowest unoccupied molecular orbital |
| | HOMO | energy of the highest occupied molecular orbital |
| | $\Delta_{LH}$ | difference between the LUMO and HOMO energies |
| | $r_{LH}$ | ratio between the LUMO and HOMO energies |
| | $q_+$ | most positive atomic charge |
| | $q_-$ | most negative atomic charge |
| | $q_{Hd}$ | charge on a possible hydrogen bond donor |
| | $q_{Ha}$ | charge on the hydrogen bond acceptor of the corresponding natural DNA |
| | $q_{5'}$ | charge on the 5′-oxygen atom |
| | $\Delta_q$ | difference between the $q_{Hd}$ and $q_{Ha}$ charges |
| | μ | dipole moment |
| | α | atomic polarizability |
| | OCC | area of weighted EEVA descriptor[a] for the occupied orbitals |
| | VIR | area of weighted EEVA descriptor[a] for the virtual orbitals |
| **II** | PHI | hydrophilic solvent-accessible surface area[b] |
| | PHO | hydrophobic solvent-accessible surface area[b] |
| | $r_{PP}$ | ratio between the PHI and PHO surfaces |
| | VM | molecular volume[b] |
| | LogP | log of the octanol/water partition coefficient[c] |
| | $M_r$ | molecular weight |

[a]Sum of overlapped Gaussian curves centered on the MO energies; for an exact definition of the EEVA descriptor see ref 14.
[b]Computed according to refs 22 and 23.
[c]Estimated partition coefficient calculated as in ref 28.

of the leave-one-out cross-validation, $r^2_{CV}$.[29] The reliability of each term is indicated by the *t*-ratio statistic, *t*-ratio, and the condition index, CI (for a given coefficient, CI is defined as the square root of the ratio between the largest eigenvalue and the eigenvalue of that particular coefficient). Good quality is indicated by large *F* and *t*-ratio values; small *s* and CI values; and $r^2$ and $r^2_{CV}$ values close to one.

Table 3 shows that the derived models account for as much as 90% of the variance in the experimental data and have reasonable predictive ability $\left(r^2_{CV} \geq 0.86\right)$. For all cellular lines $-\log(IC_{50})$ exhibits a reasonable correlation with a *two-parameter* relationship (eqs 1–3) in which the *LUMO* (the energy of the lowest unoccupied orbital) term is always present. Only in one case (cell Molt4/C8) the stepwise regression led to a three-parameter equation (eq 0b), which was nevertheless replaced by an improved two-parameter equation in the next step (eq 2b). Moreover, these regression equations have an acceptable dimension as their number of variables is 7 times less the number of observations. Apart from LUMO, the other descriptor found important for modeling the inhibitory activity is PHO (the hydrophobic solvent-accessible surface area), or LogP (the Log of the octanol/water partition coefficient), or $M_r$ (the molecular weight). Overall, the QSAR-model equations indicate that decreasing LUMO and increasing hydrophobicity (PHO or LogP) or molecular size ($M_r$) enhances the activity. Thus, the compounds' inhibitory activities on the three cellular lines seem to follow the same pattern regulation, suggesting similar mechanisms of action. Contributions from both two classes of descriptors lie around evenly, LUMO having in some cases a slightly larger weight (eqs 1a and 1c).

LUMO is a rough measure of the electron-accepting ability of a compound and, normally, reducing its value raises up that ability. It is noteworthy that a rather good correlation between the compounds' LUMOs and $-\log(IC_{50})$ is found for all cell lines: the lower the LUMO values, the higher the inhibitory activity ($r = -0.93$, $r = -0.92$ and $r = -0.93$ for L1210/0, Molt4/C8 and CEM/0, respectively). Actually, the chlorophenylpurine derivative (compound **10**) has the lowest LUMO and exhibits the highest activity. This marked dependence may be helpful in designing other more effective anticancer candidates: for example, a new candidate could be the chlorophenylpurine derivative with its chlorine atom replaced by a fluorine atom. Notice that the LUMO of this new fluorine derivative (−0.0374 hartrees) is lower than that of compound **10** (−0.0362 hartrees). To gain further insight into this problem, the LUMO of the natural DNA nucleosides deoxyadenosine, deoxythymidine and dexocytidine (the par-

**Table 3.** Most significant QSAR models for modeling the inhibitory activity on the L1210/0, Molt4/C8 and CEM/0 cell lines

| Cell line | No. | Regression equations | $r^2$ | *s* | *F* | $r^2_{CV}$ | *t*-ratio[a] CI |
|---|---|---|---|---|---|---|---|
| L1210/0 | 1a | $-\log(IC_{50}) = -55.0(\pm8.5)\text{LUMO} + 0.003(\pm0.001)\text{PHO} - 4.66(\pm0.25)$ | 0.931 | 0.171 | 88 | 0.913 | 6.5:3.9<br>7.5:15.8 |
| | 2a | $-\log(IC_{50}) = -33.2(\pm11.2)\text{LUMO} + 0.214(\pm0.047)\text{LogP} - 3.47(\pm0.16)$ | 0.943 | 0.154 | 109 | 0.917 | 3.0:4.6<br>13.4:4.2 |
| | 3a | $-\log(IC_{50}) = -41.3(\pm8.2)\text{LUMO} + 0.009(\pm0.002)M_r - 5.93(\pm0.39)$ | 0.957 | 0.135 | 144 | 0.946 | 5.1:5.5<br>31.4:7.4 |
| Molt4/C8 | 1b | $-\log(IC_{50}) = -45.5(\pm19.0)\text{LUMO} + 0.248(\pm0.079)\text{LogP} - 3.55(\pm0.27)$ | 0.899 | 0.261 | 59 | 0.863 | 2.4:3.1 |
| | 0b | $-\log(IC_{50}) = -54.0(\pm15.7)\text{LUMO} + 0.077(\pm0.092)\text{LogP} + 0.003\ \text{PHO}(\pm0.001) - 4.81(\pm0.53)$ | 0.934 | 0.211 | 62 | | 3.4:0.8:2.6<br>13.4:4.7:28.8 |
| | 2b | $-\log(IC_{50}) = -63.8(\pm10.4)\text{LUMO} + 0.004(\pm0.001)\text{PHO} - 5.17(\pm0.30)$ | 0.936 | 0.208 | 95 | 0.914 | 6.2;4.7 |
| | 3b | $-\log(IC_{50}) = -47.8(\pm11.5)\text{LUMO} + 0.012(\pm0.002)M_r - 6.80(\pm0.55)$ | 0.947 | 0.190 | 116 | 0.925 | 4.2:5.3 |
| CEM/0 | 1c | $-\log(IC_{50}) = -52.3(\pm11.3)\text{LUMO} + 0.245(\pm0.001)\text{LogP} - 3.70(\pm0.33)$ | 0.913 | 0.254 | 69 | 0.880 | 2.8:3.2 |
| | 2c | $-\log(IC_{50}) = -72.7(\pm11.3)\text{LUMO} + 0.004(\pm0.001)\text{PHO} - 5.22(\pm0.33)$ | 0.931 | 0.227 | 88 | 0.908 | 6.4:3.9 |
| | 3c | $-\log(IC_{50}) = -57.0(\pm12.4)\text{LUMO} + 0.011(\pm0.002)M_r - 6.77(\pm0.59)$ | 0.944 | 0.204 | 110 | 0.925 | 4.6:4.6 |

[a]*t*-ratio and condition index (CI) for the variable terms in the regression equations (given in the same order). CI values for the terms in eqs 1b–3b and 1c–3c are equal to those of 1a–3a.

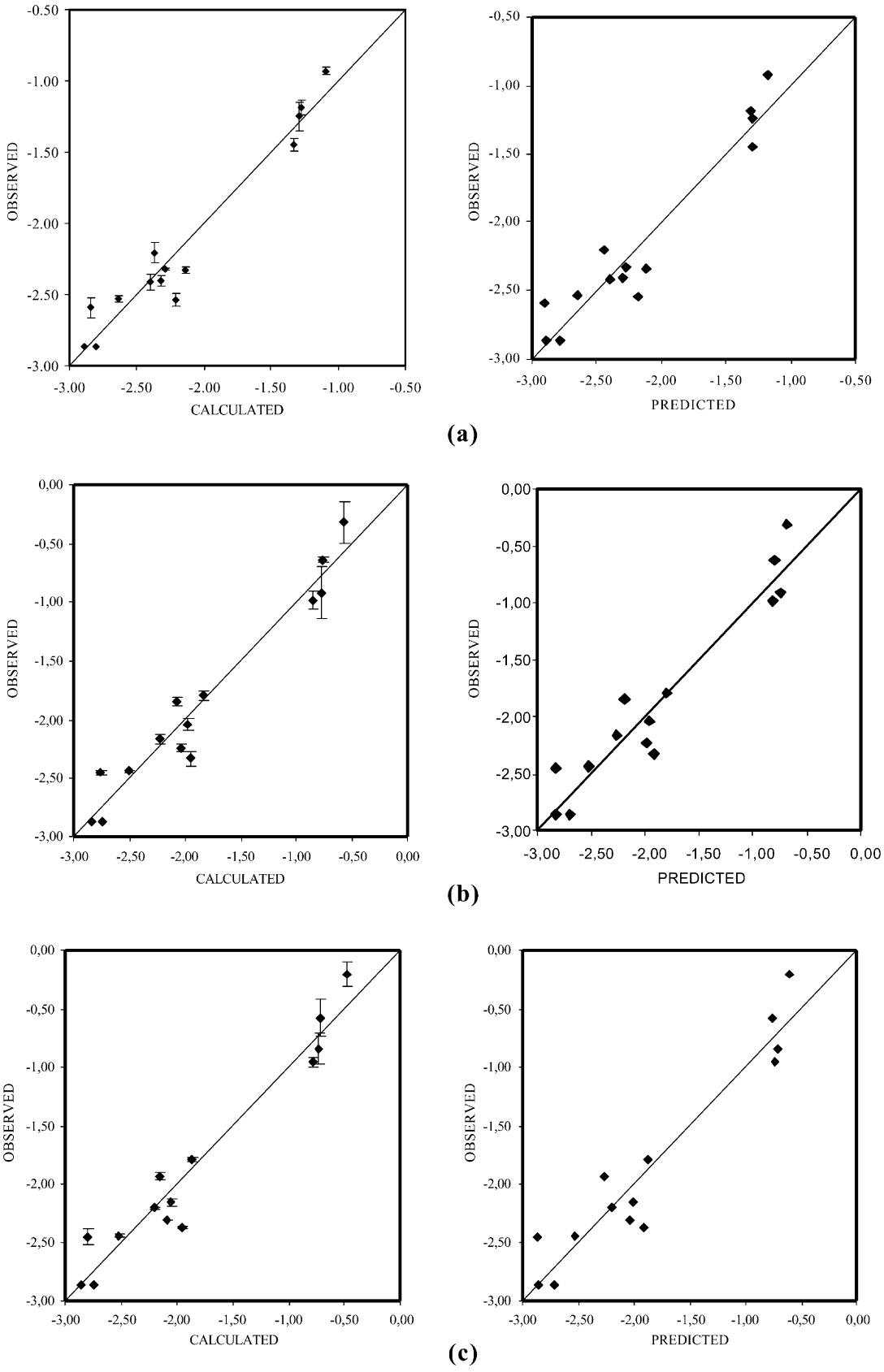



**Figure 1.** Plots of observed versus calculated (left) and observed versus predicted (right) for eqs 1a, 2b and 2c: (a) L1210/0 cells; (b) Molt4/C8 cells; (c) CEM/0 cells. The bars denote the errors of the experimental data.

ent compounds of the synthesized compounds **1**–**11**, **12**–**13** and **14**, respectively) were calculated. The fact that they have high lying LUMOs (−0.0157, −0.0155 and −0.0207 hartrees, respectively) lends support to our guessed LUMO–activity relationship.

Considering now the role of the class II descriptors, Table 3 shows that the best correlations occur for the QSAR models that include the $M_r$ descriptor (eqs 3a–3c). The $r$ values, larger $F$ statistics, and smaller $s$ values are indicative of the improved statistical significance of the models. At first sight these QSAR analyses could be interpreted such that activity was dependent upon molecular size rather than hydrophobicity. However, better reliability of the terms of such $M_r$-based models is not readily apparent ($t$-ratio values) and, especially, the fact they have a CI value greater than 30 points out to a serial problem of collinearity. Further, QSAR models of roughly the same quality and predictive ability are obtained with both PHO and LogP descriptors (eqs 1 and 2), and these being less subjected to collinearity problems (CIs $<<30$). Thus a better explanation of the inhibitory activity of these compounds is that it is strongly related also to *hydrophobicity*. Comparison of eqs 1 and 2 suggests that hydrophobicity is better parameterized with descriptor PHO than LogP. Hence, the best QSAR models for describing the present data are eqs 1a, 2b and 2c. As can be seen, for all cell lines these models are satisfactory in both statistical significance and predictive ability (Fig. 1).

It is important to know the role of hydrophobicity in modeling the compounds' activity. The results in Tables 3 and 4 clearly indicate that the more hydrophobic compounds have higher inhibitory activity with respect to all cellular lines. A careful analysis of the molecular structures shows that the markedly activity increase of compounds **8**–**11**, in comparison with compounds **1**–**7**, is certainly due to the larger hydrophobic contribution of their phenyl moiety. Furthermore, among compounds **1**–**7**, comparing **6** and **7**, compound **7** is more active than **6** for all cell lines, probably owing to its extra methyl group. For the same reason, as expected compound **9** is also more active than compound **8**. Surprisingly, the absolute hydrophobic solvent-accessible surface area rather than the portion of hydrophobic area over the total area contributes most, again suggesting that molecular size does not matter for this set of derivatives. This is also confirmed by the fact that the hydrophobic area and the molecular volume are strongly correlated ($r = 0.99$). Possibly the activity of the compounds may still be improved if the size of their hydrophobic area is enlarged.

**Table 4.** Values of selected QSAR descriptors[a] calculated for the indane derivatives

| Compd | LUMO (eV) | PHO (Å$^2$) | LogP | $M_r$ (g mol$^{-1}$) |
|---|---|---|---|---|
| **1** | −0.0179 | 385 | 1.38 | 295 |
| **2** | −0.0168 | 535 | 2.16 | 337 |
| **3** | −0.0156 | 530 | 2.24 | 337 |
| **4** | −0.0174 | 511 | 3.15 | 335 |
| **5** | −0.0224 | 380 | 0.95 | 296 |
| **6** | −0.0221 | 458 | 1.38 | 310 |
| **7** | −0.0211 | 502 | 1.89 | 324 |
| **8** | −0.0331 | 557 | 5.22 | 356 |
| **9** | −0.0326 | 586 | 5.72 | 370 |
| **10** | −0.0362 | 579 | 5.36 | 390 |
| **11** | −0.0322 | 590 | 5.32 | 386 |
| **12** | −0.0154 | 342 | 0.96 | 272 |
| **13** | −0.0144 | 378 | 1.71 | 286 |
| **14** | −0.0169 | 341 | 0.92 | 271 |

[a]Descriptors of the best linear models found (see Table 3).

Finally, some points should be remarked. Most of the descriptors employed here are subject to errors, like the quantum mechanics-derived descriptors (which describe conformationally independent gas-phase molecular properties), and this can deflate the statistical fit of the QSARs. This could be solved, for instance, by resorting to a more refined quantum method and/or conformationally dependent molecular descriptors.[30] As noted before, several descriptors were found collinear (e.g., the pairs $M_r$/PHO and $M_r$/LogP) and this might as well result from the rather poor structural variability of this set of indane derivatives. Ideally testing a larger, more diverse set of compounds could circumvent such problem.

## Concluding Remarks

In this work, 14 novel carbocyclic nucleosides indane derivatives have been synthesized and investigated thoroughly. The antitumor properties of these derivatives have been assessed by their $IC_{50}$ inhibitory patterns in the proliferation of murine leukemia and human T-lymphocyte cells. The synthesized indane derivatives have fairly strong antitumor activity considering the $IC_{50}$ values attained for the three cell lines. Moreover, they depict the same activity pattern and trend, thus suggesting similar mechanisms of action. Among them, the chlorophenylpurine derivative shows the highest activity.

On the other hand, the QSAR study enabled us to characterize the most relevant properties of the very same compounds. Based on the obtained results, it seems reasonable to conclude that the same molecular structural features are responsible for the compounds' biological activity, these being the electronaccepting ability and the hydrophobicity. The correlation equations for that activity, even though they enclose a very simple two-parameter relationship, correctly predict the experimental data observed for all cells. Further, they may help in suggesting the types of molecular modifications needed to obtain compounds of this sort with enhanced biological activity.

Naturally, there is still room for improvement in search for new potential anticancer agents; indeed, it may be too much to expect that such a simple linear approach should work for modeling the activity of chemicals beyond the present indane derivatives; and, we hope to be able to report on that shortly.

## Acknowledgements

S.-W.Y. is pleased to acknowledge the Portuguese Fundação para a Ciência e Tecnologia (FCT) for a post-doc grant

(SFRH/BPD/7125/2001). M.N.D.S.C. thanks also the Portuguese FCT for financial support that made possible her sabbatical year in Santiago de Compostela (Spain).

## References and Notes

1. De Clercq, E. *J. Med. Chem.* **1995**, *38*, 2941.
2. Huang, P.; Farquhar, D.; Plunkett, W. *J. Bio. Chem.* **1990**, *265*, 11914.
3. Bressi, J. C.; Choe, J.; Tough, M. T.; Buckner, F. S.; van Voorhis, W. C.; Verlinde, C. L. M. J.; Hol, W. G. J.; Gelb, M. H. *J. Med. Chem.* **2000**, *43*, 4135.
4. Vince, R.; Hua, M. *J. Med. Chem.* **1999**, *33*, 17.
5. Montgomery, J. A. *Med. Res. Rev.* **1982**, *2*, 271.
6. Abad, F.; Alvarez, F.; Fernandez, F.; Garcia-Mera, X.; Rodriguez-Borges, J. E. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 1127.
7. Fernandez, F.; Garcia-Mera, X.; Morales, M.; Rodriguez-Borges, J. E. *Synthesis* **2001**, *2*, 239.
8. Escobar, M.; Fernandez, F.; Garcia-Mera, X.; Rodriguez-Borges, J. E. *Nucleosides Nucleotides* **1999**, *18*, 625.
9. (a) De Clercq, E. In *In Vivo and Ex Vivo Test Systems to Rationalize Drug Design and Delivery*; Cromelin, D., Couvreur, P., Duchene D., Eds.; Editions de Sante: Paris, 1994. (b) De Clercq, E.; Descamps, J.; Verhelst, G.; Walter, R. T.; Jones, A. S.; Torrence, P. F.; Shugar, D. *J. Infect. Dis.* **1980**, *141*, 563.
10. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
11. Xiao, Z.; Xiao, Y.-D.; Feng, J.; Golbraikh, A.; Tropsha, A.; Lee, K.-H. *J. Med. Chem.* **2002**, *45*, 2294.
12. McKinney, J. D.; Darden, T.; Lyerly, M. A.; Pedersen, L. G. *Quant. Struc. Act. Relat.* **1985**, *4*, 166.
13. Waller, C. L.; McKinney, J. D. *J. Med. Chem.* **1992**, *35*, 3660.
14. Tuppurainen, K.; Ruuskanen, J. *Chemosphere* **2000**, *41*, 843.
15. Fernandez, F.; Garcia-Mera, X.; Morales, M.; Rodriguez-Borges, J. E.; De Clercq, E. *Synthesis* **2002**, *8*, 1084.
16. (a) Mitsunobu, O.; Wata, M.; Sano, T. *J. Am. Chem. Soc.* **1972**, *94*, 679. (b) Mitsunobu, O. *Synthesis* **1981**, 1. (c) Girard, F.; Lee, M. G.; Agrofoglio, L. A. *J. Heterocycl. Chem.* **1998**, *35*, 911.
17. Suzuki, A. *J. Organomet. Chem.* **1999**, *576*, 147, and references therein.
18. Unpublished results.
19. *The QUANTA® Program*; Molecular Simulations Inc.: San Diego, 1999.
20. Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 209.
21. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakreswki, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuk, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowki, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98,* revision A.11; Gaussian Inc.: Pittsburgh, PA, 1998.
22. Connolly, M. L. *J. Appl. Cryst.* **1983**, *16*, 548.
23. Connolly, M. L. *Science* **1983**, *221*, 709.
24. *SPSS for Windows*, Rel. 11.5.1; SPSS Inc.: Chicago, 2002.
25. Ford, M. G.; Livingstone, D. J. *Quant. Struct. Act. Relat.* **1990**, *9*, 107.
26. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
27. Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.
28. Parham, M.; Hall, L.; Kier, L. *Accurate Prediction of LogP using E-state Indices with Neural Network Analysis*, 220th ACS National Meeting, Washington, DC, 2000.
29. Schaper, K. J. *Quant. Struct. Act. Relat.* **1999**, *18*, 354.
30. Mekenyam, O.; Ivanov, J.; Karabunarliev, S.; Bradbury, S. P.; Ankley, G. T.; Karcher, W. *Environ. Sci. Technol.* **1997**, *31*, 3702.