Nonequivalent Effects of Diverse Log*P* Algorithms in Three QSAR Studies

Eduardo Borges de Melo[a, b]* **and Márcia Miguel Castro Ferreira**[b]

[a] Curso de Farmácia, Centro de Ciências Médicas e Farmacêuticas, Universidade Estadual do Oeste do Paraná – Unioeste, Rua Universitária, 2069, 85819-110, Cascavel, Paraná, Brazil.

[b] Theoretical and Applied Chemometrics Laboratory (http://lqta.iqm.unicamp.br), Institute of Chemistry, University of Campinas – Unicamp, Campinas, São Paulo, 13083-970, Brazil

*e-mail: ebmelo@unioeste.br

Keywords: Log*P*, QSAR, HIV inhibitors, QSAR validation, Structure-activity relationships, Drug design

Received: August 16, 2008; Accepted: June 11, 2009

DOI: 10.1002/qsar.200810125

Abstract

Despite of the availability and facility of accessing several algorithms for calculation of Log*P* in QSA(P)R studies, articles typically do not describe the selection procedure for the method used. Therefore, three studies to verify the influence of different Log*P* algorithms on building QSAR models were performed. Two QSAR data sets from the literature (forty-two tricyclic phtalimide inhibitors of HIV-integrase and fourty-six TIBO derivatives inhibitors of HIV-reverse transcriptase) were used together with Log*P* calculated by thirteen algorithms, and several regression models were constructed and compared. A new QSAR study for 4,5-dihydroxypyrimidine carboxamides inhibitors of HIV-1 integrase was also performed. The explained and predicted variance, results from external validation, leave-*N*-out cross-validation and **y**-randomization test were analyzed for all models from the three data sets. Despite the same physicochemical meaning, Log*P*'s calculated by distinct methods may show different levels of contribution to the model. This observation comes out from the comparison of validated models. These results indicate that the arbitrary choice of one specific algorithm for Log*P* calculation, as is usual in QSA(P)R studies , does not necessarily lead to the highest quality model for the analyzed data set.

1 Introduction

Parameters that encode physicochemical and molecular properties, generally designated as molecular descriptors, are used in quantitative structure-activity (or property) relationships studies, QSA(P)R. The descriptors are employed for building quantitative (mathematical) models to analyze correlation between the chemical structure and specific biological activity or property. Of particular value are the descriptors that encode information about the drugs transport and drug-receptor binding [1].

1-Octanol/water partition coefficient (*P*) is certainly one of the most important among thousands currently available descriptors, being defined as the concentration ratio of a substance in the organic and aqueous phases of a two-compartment system under equilibrium conditions [1]. Many biological processes, such as biomembrane-mediated passage of a drug from blood (an aqueous media) to tissues depend on the partition coefficient [2]. Due to theoretical reasons and the fact that values of *P* can vary by 12 orders of magnitude (from 10^{-4} to more than 10^{8}), commonly the logarithm (Log*P*) is used to characterize this property [3–5].

Supporting information for this article is available on the WWW under www.qcs.wiley-vch.de

Abbreviations: ***ARE*$_{pred}$** average relative error of prediction; **ES** external validation set; **HIV** human immunodeficiency virus; **LNO** leave-*N*-out crossvalidation; **PHYSPROP** Physical Properties Database; **PLS** Partial Least Squares; ***PRESS*$_{cal}$** predictive residual sum of squares of calibration; ***PRESS*$_{val}$** predictive residual sum of squares of calibration of validation; **Q^2_{LNO}** correlation coefficient of leave-*N*-out cross-validation; **Q^2_{LOO}** correlation coefficient of leave-one-out crossvalidation; **QSAR** quantitative structure-activity relationship; **R^2** correlation coefficient of calibration; **R^2_{pred}** correlation coefficient of prediction; ***SEC*** standard error of calibration; ***SEP*** standard error of prediction; ***SEV*** standard error of cross-validation; ***SSy*** sum of squares of the response values; **TIBO** tetrahydroimidazo[4,5,1-jk][1, 4]benzodiazepinone; **TS** training set.

Besides of being involved in the pharmacokinetic phenomena, Log*P* can also be related to the drug/receptor interactions [6]. The determination of Log*P* can be helpful to a better understanding of how this property is associated with the hydrophobic interactions and the phenomenon of entropy-enthalpy compensation, which is related to solvation/desolvation processes [7].

Log*P* is widely used in obtaining models for the prediction of molecular behavior in pharmaceutical, environmental, biochemical and toxicological sciences since it is a good measure of molecular lipophilicity [3,8,9]. The main methodology to determine *P* is based on the assessment of the relative distribution of a substance in a biphasic system formed by 1-octanol/aqueous buffer under agitation ('Shake-Flask') [6,10,11], however, other approaches are also available [12–14].

The use of experimental values of Log*P* as a descriptor can provide more realistic models in QSA(P)R studies. However, experimental determination of Log*P* can be a laborious, time consuming and an expensive procedure. Such situation, and the existence of vast amount of new natural or synthesized molecules are quite problematic factors for databases, as THOR [15] or PHYSPROP [16], which have to remain constantly updated. Thus, computational approaches are currently very valuable tools to derive Log*P*'s from chemical structures in QSA(P)R studies.

The first way to derive Log*P*'s from chemical structures was the π-system [17–19]. Actually, various algorithms with this objective, commercial or freeware, are available [20]. Two principal approaches for Log*P* calculation are used: (a) the substructure method based on fragments or atoms (or both), and (b) the whole molecule method, which is based on molecular properties [18]. Studies performed for distinct sets of compounds and including a comparison between experimental and predicted Log*P* values by using different algorithms, have shown that there is no unique algorithm that assures the best prediction of Log*P*, despite of the fact that all calculated Log*P* values have the same physical meaning [6, 9, 21–23]. Overall, good agreement among calculated Log*P*'s has been observed by Karthikeyan and co-workers [24] for a large set of drugs, but this does not imply the same trend for a specific set of compounds.

Even though, current QSA(P)R studies do not specify how and why an algorithm was selected arbitrarily for performing a Log*P* calculation. However, regarding the ease of accessing several algorithms for Log*P* generation, is it acceptable to build a model using a specific algorithm without testing for others? Is it not possible that better results could be obtained if the algorithm 'B' is used instead of 'A', leading to a more robust model? In a previous study by Ferreira and Kiralj [25] it has been shown that various algorithms for Log*P* calculation result in values encoding different structural information and this, consequently, lead to different QSAR models. In this work, the relevance of the algorithm selection for performing Log*P* calculation in QSA(P)R studies is revised, extended and better explored. For this purpose, three data sets were used to test sixteen distinct algorithms.

2 Methods

2.1 Data Sets

Three sets of anti-HIV compounds with no experimental values of Log*P* were selected from the literature [25–27]. The basic structures of the compounds are presented in Figure 1 and all the molecular structures are available in the Supplementary Material, Figures S1–S3.

The first two data sets were previously utilized by Bansal and co-workers [26] and Huuskonen [27] in 2D-QSAR studies, where a Log*P* descriptor was included in both published models. The two data sets were selected with the objective to evaluate the influence of liphophilicity (Log*P*) calculated by distinct algorithms on the final model. The data sets were split into training sets (TS) and test sets used in external validation (ES). TS1 [26] is constituted by forty-two tricyclic phtalimide analogues reported as HIV-1 integrase (HIV-1 IN) inhibitors, with biological activity expressed as pIC_{50} ($-\log IC_{50}$). The original model was built on a set of thirty compounds and included descriptors MLog*P*, RBF (rotable bond fraction), nPhX (number of halogen atoms bonded to carbon atoms in the aromatic ring) and Jhete (Balaban-type index derived from electronegativity-weighted distance matrix) (Supplementary Material, Table S1).

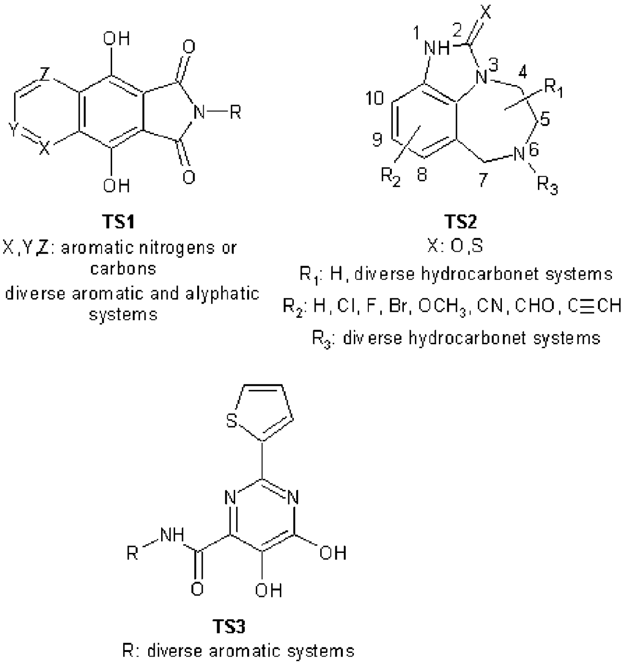
OH
O
Z
N–R
Y
X
OH
O
TS1
X,Y,Z: aromatic nitrogens or carbons
R: diverse aromatic and alyphatic systems
X
1
2
NH
3
N
4
R1
10
5
9
N
6
R2
8
7
R3
TS2
X: O,S
R1: H, diverse hydrocarbonet systems
R2: H, Cl, F, Br, OCH3, CN, CHO, C≡CH
R3: diverse hydrocarbonet systems
S
N
N
R
NH
OH
O
OH
TS3
R: diverse aromatic systems

Figure 1. Basic structures of the selected training sets (TS).

TS2 [27] is constituted by forty-six tetrahydroimidazo[4,5,1-jk][1, 4]benzodiazepinone (TIBO) derivatives reported as HIV-1 reverse transcriptase (HIV-1 RT) inhibitors, where the biological activities were measured and expressed as Log 1/C (C is the IC_{50}, the effective concentration of a compound to achieve 50% protection of MT-4 cell against the cytopathic effect of HIV-1). The original model was built on forty-one compounds and five selected descriptors: CLog*P* and atom-level E-state indices for atoms C2, C4, C8 and C9 (Supplementary Material, Table S2).

TS3 [28] is a set of thirty-three 4,5-dihydroxypyrimidine carboxamides reported as HIV-1 IN inhibitors. An original QSAR study with Log*P* in the model was developed and is presented in this work. This TS is interesting since lipophilicity might be important for the inhibitory potency, because the possibility of the interaction between the aromatic side and an apolar environment located in the HIV-1 IN active site [28, 29]. The biological activity was expressed in terms of the necessary concentration for 50% of inhibition of the strand transfer reaction (IC_{50}, in nM), and transformed to pIC_{50}.

2.2 Methods for LogP Calculation

The algorithms for Log*P* calculation employed were the following: freeware methods ALOG*P*s, AB/Log*P*, ACLog*P*, ALog*P*, COSMOfrag, miLog*P*, MLog*P*, XLog*P*2, XLog*P*3, KOWWIN (available on-line at www.vcclab.org), molLog*P* (available on-line at www.molsoft.com) and IA-Log*P* (previously available on-line at www.logp.com, but no any longer), a demo version of CSLog*P* (available on-line at www.chemsilico.com), a freeware version of ACD/Log*P* (commercial and freeware versions available at www.acdlabs.com/download/logp.html) and the commercial packages ChemOffLog*P* and CLog*P* (versions implemented in Chem 3D Ultra 5.0). IALog*P* was used only for TS3 because it was not available on-line at the time when TS1 and TS2 were included in this work. Complete information about the methods and the values obtained for each sample (neutral chemical structures) can be found in the Supplementary Material, Tables S3 – S6.

2.3 QSAR Studies

For TS1 and TS2 studies, the data matrices of dimensions (30×4) and (41×5), respectively, were extracted from the literature. For each compound, the Log*P* values were calculated by the other fourteen algorithms (all values are available in the Supplementary Material, Tables S4 and S5, respectively) and new fourteen matrices were built, differing only in the Log*P* values. Matlab 7 software [30] was used to build the models with multiple linear regression (MLR), the regression method used in the original works.

For TS3, values of Log*P* descriptors were calculated using the sixteen algorithms available at that time (Supplementary Material, Table S6). Initially, the Pearson correlation coefficient (r) between each Log*P* and pIC_{50} was calculated, and the algorithm with highest r was selected and added to other 161 calculated molecular descriptors obtained by several software (Supplementary Material, Table S7). An a priori variable selection was performed and the descriptors with $|r| < 0.3$ were eliminated considering that they did not contain relevant information. Thus, the training set was reduced to 63 descriptors. The models were built using Partial Least Squares (PLS) regression [31] implemented in the Pirouette 4 software [32], on the data previously autoscaled. The final descriptors were selected by means of the most significant PLS regression coefficients. The compounds with Studentized residuals above 2σ were considered outliers. The data matrix for this QSAR model, with dimensions (29×4), was used to build the other 15 matrices by substituting the Log*P* descriptors, similarly to TS1 and TS2. PLS models were built for all these matrices.

According to literature [33 – 39], rigorous validation procedures are necessary to assure statistical reliability of the QSAR models. This approach has adopted in this work. For all models, leave-one-out cross-validation (LOO) was applied to determine the correlation coefficient of cross-validation, Q^2_{LOO} (Supplementary Material, Table S8). The correlation coefficient of calibration, R^2, was also calculated, as a measure of quality of fitting. The recommended limits for these parameters are $R^2 \geq 0.6$ and $Q^2_{\mathrm{LOO}} \geq 0.5$ [35, 38]. The corresponding errors *SEC* and *SEV* should be as smaller as possible. The $PRESS_{\mathrm{val}}$ values should be smaller than the sum of squares of the response values (SS_Y) [30].

The tabulated *critical-F* ($F_{p,n\text{-}p\text{-}1}$) values, or *cF*, where n is the number of compounds and p is the number of descriptors or latent variables in the final model, were obtained for each TS and compared with the result obtained from the *F-test* ($\alpha = 0.05$). For this test, the higher the difference between the *cF* and the *F-test* value, the more statistically significant is the model [40].

For the external validation, the external sets ES1 and ES2, the same used by Bansal and co-workers [26] and Huuskonen [27], consisted of eight and twenty-four compounds, respectively. The ES3, corresponding to TS3, contained seven compounds and is considered appropriate because the data split follows literature recommendations [35], being a significant sample of the training set (24%, without outliers).

The robustness of the models were examined by leave-*N*-out cross validation (LNO, with $N = 1$ to 10 for TS1 and TS2, and $N = 1$ to 6 for TS3). The presence of chance correlation was checked by the **y**-randomization test [38]. Robustness is a measure of internal performance which shows whether the model is not significantly affected by small and deliberate changes in their parameters [42], as in the LNO cross-validation. Chance correlation in QSAR means that any variable which is not in reality related to the drug action can be well statistically correlated with biological activity, what results in statistically acceptable but nonsense models, and can be accessed by the **y**-randomization test. Both strategies can compare the influence of the different Log*P* descriptors in the models, because they can show if any of them leads to an unreliable model.

The LNO cross-validation employs smaller training sets than the LOO procedure, and QSAR models with a high Q^2_{LNO} and stable can be considered robust [43]. For each value of *N*, the pre-randomization of all rows of the data (**X** with corresponding **y**) was performed three times in order to decrease the impact that the withdrawal of sets of samples in specific sequences could have on the values of Q^2_{LNO}. It is expected that the average values obtained from the triplicate tests are close to that of Q^2_{LOO}, with small standard deviations [35]. The **y**-randomization test was performed ten times [33]. The adopted limits are based on the intercepts values proposed by Eriksson and co-workers [39], but in this work all randomized models should present $R^2 \leq 0.3$ and $Q^2_{\mathrm{LOO}} \leq 0.05$. The **y**-randomization test is useful to verify the possibility that models with high values of R^2 and Q^2_{LOO} could suffer from chance correlation [34]. LNO and **y**-randomization tests were performed in Matlab 7 [30] and the plots were built in the DataFit 9 [42].

Taking into account the specific objectives of this work, the interpretation of the models was not considered relevant. The interpretations of the original models for TS1 and TS2 can be found in the literature [26, 27].

3 Results and Discussion

3.1 Analysis for TS1 and TS2

The first step in TS1 and TS2 studies was to check if any other Log*P* descriptor would have a *r* value higher than that from Log*P* used in the original works [26, 27]. The results obtained for the two training sets are in Table 1.

For TS1, correlation coefficients vary from $r = 0.61$ for COSMOfrag to $r = 0.24$ for Clog*P*. The descriptor used in the literature was MLog*P*, with $r = 0.51$. In the case of TS2, the highest *r* value obtained was $r = 0.54$ for ACLog*P* and the lowest $r = 0.05$ for CSLog*P*, while the CLog*P* descriptor presented $r = 0.30$, a difference of 0.24 with respect to ACLog*P*.

It is interesting to note that in both cases, the literature descriptor did not yield the highest value of *r*. For instance, the CLog*P* descriptor had the second lowest *r* value in TS2, being higher only from that of CSLog*P*. Similar result were obtained for TS1, where MLog*P* descriptor also possessed the second lowest *r* value. These preliminary results indicate that it might not be enough simply to select the Log*P* descriptor with the highest *r* to the biological activity in a QSAR study.

Multivariate models for TS1 and TS2 were obtained (Supplementary Material, Tables S9 and S10) and compared for their internal and external statistical quality (Tables 2 and 3), and validated for their robustness and possible chance correlation using the LNO cross-validation and **y**-randomization tests, respectively.

Table 1. Pearson correlation coefficient (*r*) between the algorithms and the biological activities of TS1 and 2.

Algorithm	*r* TS1	*r* TS2
AB/Log*P*	0.52	0.36
ACD/Log*P*	0.55	0.45
ACLog*P*	0.51	0.54
ALog*P*	0.60	0.46
ALOG*P*s	0.57	0.32
ChemOffLog*P*	0.59	0.52
CLog*P*	0.24	0.30 [a]
COSMOfrag	0.61	0.43
CSLog*P*	0.54	0.05
KOWWIN	0.58	0.38
miLog*P*	0.57	0.41
MLog*P*	0.51 [a]	0.43
molLog*P*	0.32	0.37
XLog*P*2	0.55	0.31
XLog*P*3	0.53	0.39

[a] values avaliables in the references.

3.2 Analysis of QSAR Models for TS1

Evaluating all statistical parameters (Table 2), the CSLog*P* model can be considered as the most appropriated model for TS1. This model is equivalent to that from the literature. Although the statistical quality of the MLog*P* model could be considered superior with respect to the parameters R^2, $PRESS_{\mathrm{cal}}$, Q^2_{LOO}, $PRESS_{\mathrm{val}}$ and *F*, the differences between the two models are very small. However, the model CSLog*P*, besides being equivalent to MLog*P*, has a relatively low ARE_{pred} and high R^2_{pred}. The external validation is a very important step in QSA(P)R studies and, therefore, was also considered an important step to evaluate the predictability of a model, before applying it to unknown samples. Several authors argue that only models internally and externally validated may be considered statically realistic and applicable for practical purposes [35–37]. Thus, even with the internal quality of the MLog*P* model being equivalent to CSLog*P* model, the last may be considered better due to its performance in external validation.

Both models were satisfactory in the LNO cross-validation and **y**-randomization test (Fig. 2). For the CSLog*P* model, the average Q^2_{LNO} is 0.66, the same for Q^2_{LOO}, and the standard deviations for each number of excluded samples, *N*, can be considered acceptable, (maximum deviation is 0.07 for L9O). For the model MLog*P*, the Q^2 statistics is similar (average Q^2_{LNO} is 0.65 and Q^2_{LOO} is 0.66), but much larger standard deviations are observed (see Fig. 2). In the **y**-randomization test, all values for R^2 and Q^2_{LOO} for both models are below the acceptable limits. Even so, the

Table 2. Results for explained and predicted variance and external validation of the models for TS1.

	R^2	SEC	$PRESS_{cal}$	Q^2_{LOO}	SEV	$PRESS_{val}$ [a]	F [b]	R^2_{pred}	SEP	ARE_{pred} (%)
MLog*P*[c]	0.81	0.25	1.55	0.66	0.31	2.80	27.20	0.62	0.34	5.04
AB/Log*P*	0.80	0.26	1.73	0.57	0.35	3.67	24.74	0.59	0.36	4.97
ACD/Log*P*	0.78	0.26	1.74	0.54	0.36	3.91	24.56	0.59	0.36	5.02
ACLog*P*	0.81	0.25	1.58	0.63	0.33	3.21	27.60	0.59	0.36	4.97
ALog*P*	0.82	0.25	1.53	0.65	0.31	2.99	28.85	0.31	0.46	5.98
ALOG*P*s	0.84	0.27	1.40	0.69	0.31	2.84	31.96	0.52	0.38	5.27
ChemOffLog*P*	0.82	0.25	1.57	0.65	0.31	2.98	27.97	0.50	0.40	5.63
CLog*P*	0.75	0.29	2.18	0.50	0.38	4.26	18.36	0.66	0.33	4.66
COSMOfrag	0.79	0.27	1.79	0.57	0.35	3.67	23.64	0.51	0.39	5.21
CSLog*P*	0.81	0.25	1.59	0.65	0.31	2.94	27.43	0.64	0.34	4.88
KOWWIN	0.80	0.26	1.71	0.61	0.33	3.37	25.14	0.62	0.35	4.75
miLog*P*	0.81	0.25	1.61	0.62	0.33	3.24	27.04	0.57	0.36	5.03
molLog*P*	0.78	0.28	1.92	0.44	0.40	4.76	21.70	0.56	0.37	5.48
XLog*P*2	0.80	0.26	1.68	0.63	0.32	3.17	25.58	0.55	0.37	5.31
XLog*P*3	0.81	0.25	1.59	0.65	0.31	2.98	27.41	0.52	0.38	5.27

[a] $SS_Y=8.58$; [b] $F_{4,26}=2.74$ ($\alpha=0.05$); [c] literature model.

Table 3. Results for explained and predicted variance and external validation of the models for Statiscs TS2.

	R^2	SEC	$PRESS_{cal}$	Q^2_{LOO}	SEV	$PRESS_{val}$ [a]	F [b]	R^2_{pred}	SEP	ARE_{pred} (%)
CLog*P* [c]	0.85	0.52	9.65	0.80	0.56	13.04	40.01	0.82	0.64	8.23
AB/Log*P*	0.81	0.59	12.21	0.73	0.66	17.75	30.18	0.79	0.69	10.17
ACD/Log*P*	0.83	0.56	10.93	0.76	0.61	15.50	34.53	0.83	0.63	9.05
ACLog*P*	0.82	0.57	11.29	0.76	0.62	15.72	33.18	0.82	0.65	9.14
ALog*P*	0.82	0.57	11.54	0.75	0.63	16.34	32.31	0.81	0.67	9.51
ALOG*P*s	0.80	0.61	13.13	0.72	0.66	17.95	27.57	0.75	0.77	11.06
ChemOffLog*P*	0.81	0.59	12.02	0.74	0.64	16.99	30.73	0.78	0.71	10.39
COSMOfrag	0.83	0.56	11.08	0.76	0.61	15.49	33.97	0.81	0.66	9.23
CSLog*P*	0.80	0.57	13.22	0.72	0.67	18.35	27.32	0.73	0.79	11.36
KOWWIN	0.84	0.54	10.17	0.78	0.59	14.41	37.60	0.83	0.62	8.79
miLog*P*	0.83	0.50	10.79	0.76	0.61	15.26	35.04	0.83	0.62	8.61
MLog*P*	0.82	0.58	11.78	0.74	0.64	16.97	31.52	0.81	0.66	9.46
molLog*P*	0.84	0.50	10.45	0.78	0.59	14.10	36.43	0.82	0.64	9.46
XLog*P*2	0.82	0.58	11.93	0.74	0.64	16.76	31.04	0.80	0.69	9.66
XLog*P*3	0.83	0.54	10.92	0.76	0.61	15.41	34.55	0.83	0.63	8.81

[a] $SS_Y=64.82$; [b] $F_{5,35}=2.48$ ($\alpha=0.05$); [c] literature model.

result of LNO validation is slightly superior for the model CSLog*P* showing that this model is more robust in relation to the MLog*P* model. The obtained results for all models for TS1 are available in the Supplementary Material, Tables S11 and S12.

In the case of a comparison between the *r*'s from CSLog*P* and MLog*P*, the former would have been chosen. Considering the similarities between the two models, it can be suggested that there would be a good chance of obtaining of a model formed by the same four descriptors if Bansal and co-workers [26] had used the CSLog*P* algorithm. However, in a work where several algorithms would be used, only comparing the *r*'s may not be enough to choose an algorithm.

The models obtained with the ALOG*P*s and CLog*P* descriptors also deserve a special attention. Among all models from Table 2, the former was better in R^2, Q^2_{LOO} and in the *F*-test, and the latter in the external validation. But the model CLog*P* had the worst results in the LNO cross-validation (Fig. 3). Considerable variation of the Q^2_{LNO} values with respect to the Q^2_{LOO} may be observed, and also large standard deviations at high *N* can be noticed. Besides that, unacceptable average values of Q^2_{LNO} below 0.5 occur in 50% of the cases, indicating that this model does not provide the adequate robustness and can be considered as the worst model.

The model ALog*P* deserves attention because it has the second highest *r* (Table 1). This model had explained and predicted variances equivalent to those for the models MLog*P* and CSLog*P*, but was rejected due to poor results from external validation, ($R^2_{pred}=0.31$, the only value below 0.5). Moreover, the results of the **y**-randomization test

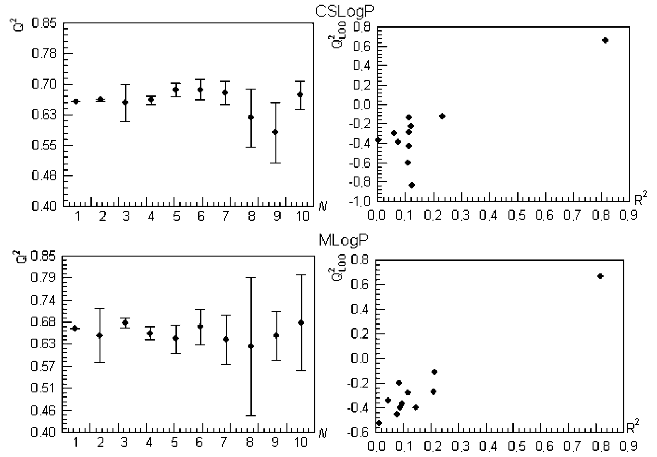

Figure 2. Plots for LNO cross-validation (left) and **y**-randomization test (right) for the models CSLog*P* and MLog*P* of the TS1.

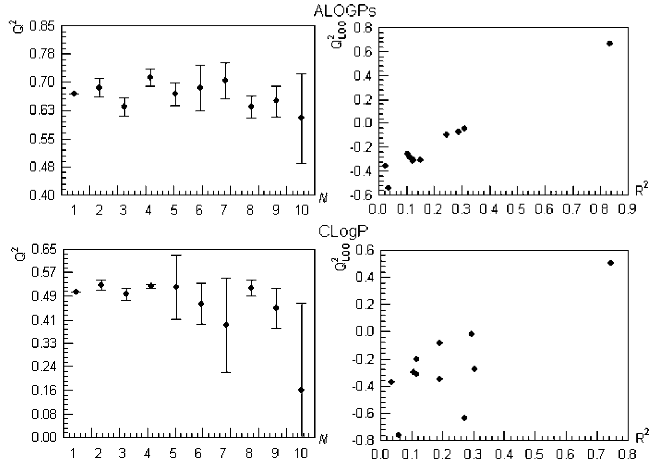

Figure 3. Plots for LNO cross-validation (left) and **y**-randomization test (right) for the models ALOG*P*s and CLog*P* of theTS1.

(Supporting Information, Table S11) clearly indicate chance correlation. In the LNO cross-validation (Supporting Information, Table S12), the result for L10O (0.47) is below the acceptable limit and low compared to Q^2_{LOO} (0.62). Therefore, this model may be considered as of the lowest quality from TS1.

3.3 Analysis of QSAR Models for TS2

The literature model with CLog*P* descriptor showed to be of the highest quality for TS2 (Table 3). KOWWIN descriptor yielded a statistically equivalent model, and the two models differed in the result of the *F*-test (40.01 for CLog*P* and 37.60 for KOWWIN), For both models, LNO validation presented average Q^2_{LNO} practically identical to Q^2_{LOO} (0.79 and 0.80 for CLog*P*, and 0.77 and 0.78 for KOWWIN), as well as maximum standard deviations of 0.04 from L10O for CLog*P*, and 0.03 from L8O for KOWWIN. The results of the **y**-randomization test have shown that none of the models possess chance correlation (Fig. 4).

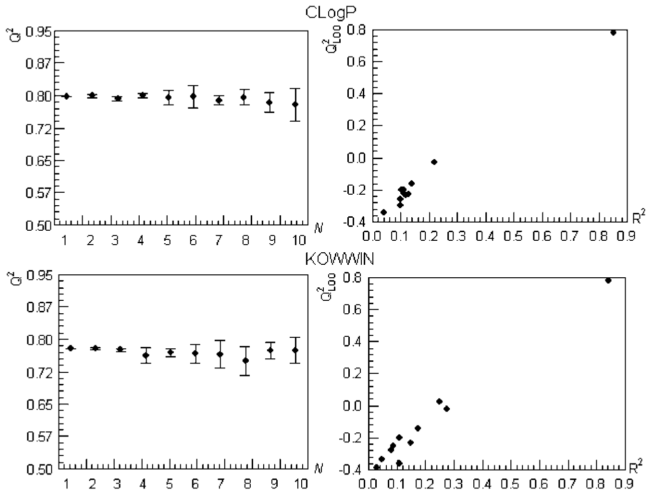

Figure 4. Plots for LNO cross-validation (left) and **y**-randomization test (right) for the models CLog*P* and KOWWIN of TS2.

The basic statistics (R^2 and Q^2_{LOO}) for other models obtained from TS2, the results of external validation and LNO cross-validation, are inside acceptable limits. However, miLog*P* model is the only one that overcomes modestly the limits for **y**-randomization test in one out of 10 randomizations (Supporting Information, Table S13).

The obtained results for TS1 and TS2 have shown that, despite that fact that the literature models are of good quality, there is a possibility to obtain improved, equivalent or even inferior quality models when other Log*P* descriptors are used. It can be concluded that not only the basic statistical parameters R^2, Q^2_{LOO} and *F*-ratio are enough to test the quality of the models, but other validations should be considered, such as external validation, LNO and **y**-randomization. Thanks to these tests, it was possible electing the model CSLog*P* as the best for TS1, and especially miLog*P* as the worst for TS2. The results of **y**-randomization test and LNO cross-validation of TS2 are available in the Supplementary Material, Tables S13 and S14.

The most important observation is that using only *r* is not enough to identify the most appropriate Log*P* descriptor to be used in a QSA(P)R study.

3.4 New Study – TS3

Similarly to the results for TS1 and TS2 (Table 1), it is possible to observe a large variation in *r* between the sixteen Log*P*'s and biological activity, for the complete data set 3 (Table 4). The results show that the best descriptors are COSMOfrag ($r=0.55$) and XLog*P*3 ($r=0.54$), and the worst are MLog*P* ($r=0.27$) and Clog*P* ($r=0.31$). Thus, the

Table 4. Pearson correlation coefficient (r) between the algorithms and the biological activities of TS3.

Algorithm	r
AB/Log*P*	0.42
ACD/Log*P*	0.47
AcLog*P*	0.42
ALog*P*	0.44
ALOG*P*s	0.47
ChemOffLog*P*	0.45
CLog*P*	0.31
COSMOfrag	0.55
CSLog*P*	0.46
IALog*P*	0.41
KOWWIN	0.43
miLog*P*	0.39
MLog*P*	0.27
molLog*P*	0.37
XLog*P*2	0.33
XLog*P*3	0.54

COSMOfrag descriptor was used to obtain the initial model and to split the data set into TS3 and ES3.

Four compounds, **8**, **11**, **34** and **39** presented Studentized residuals above 2σ, and were considered as outliers. The initial model with COSMOfrag descriptor was obtained when also using the energy of lowest unoccupied molecular orbital (LUMO), solvation connectivity index chi-0 (X0sol), and bond-type E-state index SeaC2C2aa (Supplementary Material, Table S15). This model was built by PLS regression with two latent variables.

An ES containing seven compounds (**10**, **12**, **16**, **17**, **25**, **26** and **35**), with low leverage in the model on the complete data set, was selected. The samples are good representatives of the whole pIC_{50} range and the training sets structural diversity. After this step, other 15 models were built exchanging the Log*P* descriptors as before.

Table 5 presents the statistics for the sixteen models obtained. The regression coefficients are in the Supplementary Material, Table S16. The model using COSMOfrag descriptor has the highest Q^2_{LOO} and R^2_{pred}. Besides the significant variation in r, all the models have R^2 around 0.60, indicating the importance of other descriptors to the models.

Despite that all the models were built with twenty-two samples and two latent variables, there is a reasonable variation in the amount of information contained in each model, with maximum difference of 25.75% (between COSMOfrag and CLog*P* models). Considering that LUMO, X0sol, and SeaC2C2aa descriptors are in common for all 16 models, contributions of the different Log*P* descriptors are clear. Only two Log*P*'s led to acceptable models: COSMOfrag and XLog*P*3. These models explained 64.0% and 62.0%, and predicted 52.0%, and 50.0% of total variance, respectively. They also presented the smallest values of *SEV*. The information retrieved from two latent variables was highly significant, indicating that the models used most of the available information in the original descriptors. This can explain the statistical significance observed by the high *F*-test value with respect to the *critical-F*, 3.52 (for $p=2$ and n-p-$1=19$), and also by the $PRESS_{\mathrm{val}}$ values, which are lesser than 18.46 (the result found for the SS_Y) [33]. These two models were also able to provide the best results in the external validation of this training set, with R^2_{pred} above 0.50, ARE_{pred} below 10.00%, and the lowest *SEP*.

The models COSMOfrag and XLog*P*3 show good results for LNO validation (Fig. 5), and this may be considered as the most important information in this study. There was a satisfactory performance in the **y**-randomization test only these two models, while the others present some results out of the adopted limits ($R^2 \leq 0.3$ and $Q^2_{\mathrm{LOO}} \leq 0.05$ for all results). In this case, exactly these two algorithms

Table 5. Results for explained and predicted variance and external validation of the models for TS3.

	R^2	*SEC*	$PRESS_{\mathrm{cal}}$	Q^2_{LOO}	*SEV*	$PRESS_{\mathrm{val}}$ [a]	*F* [b]	R^2_{pred}	*SEP*	ARE_{pred} (%)
AB/Log*P*	0.64	0.60	6.73	0.42	0.70	10.63	16.55	0.54	0.75	10.53
ACD/Log*P*	0.62	0.61	7.09	0.40	0.71	11.13	15.24	0.20	0.94	13.09
ACLog*P*	0.61	0.62	7.25	0.37	0.73	11.64	14.70	0.25	0.90	12.27
ALog*P*	0.61	0.62	7.25	0.37	0.73	11.59	14.68	0.28	0.88	11.94
ALOG*P*s	0.61	0.62	7.20	0.38	0.72	11.47	14.86	0.22	0.92	12.64
ChemOffLog*P*	0.63	0.60	6.91	0.41	0.70	10.84	15.89	0.04	1.08	14.88
CLog*P*	0.66	0.58	6.35	0.47	0.67	9.74	18.10	−0.09	1.19	17.56
COSMOfrag	0.65	0.55	6.46	0.52	0.61	8.94	19.52	0.59	0.68	8.96
CSLog*P*	0.61	0.62	7.24	0.44	0.68	10.26	14.73	0.34	0.82	11.19
IALog*P*	0.61	0.62	7.26	0.44	0.69	10.40	14.67	0.39	0.78	10.52
KOWWIN	0.63	0.60	6.91	0.41	0.70	10.84	15.89	0.04	1.08	14.88
miLog*P*	0.61	0.62	7.29	0.35	0.74	11.95	14.55	0.30	0.86	11.62
MLog*P*	0.61	0.62	7.29	0.41	0.70	10.89	14.57	0.32	0.85	11.34
molLog*P*	0.61	0.62	7.20	0.38	0.72	11.49	14.84	0.23	0.93	12.76
XLog*P*2	0.60	0.62	7.34	0.37	0.73	11.69	14.38	0.43	0.80	11.04
XLog*P*3	0.64	0.59	6.57	0.50	0.65	9.23	17.21	0.55	0.70	9.37

[a] $SS_Y = 18.46$; [b] $F_{2,19} = 3.52$ ($\alpha = 0.05$).

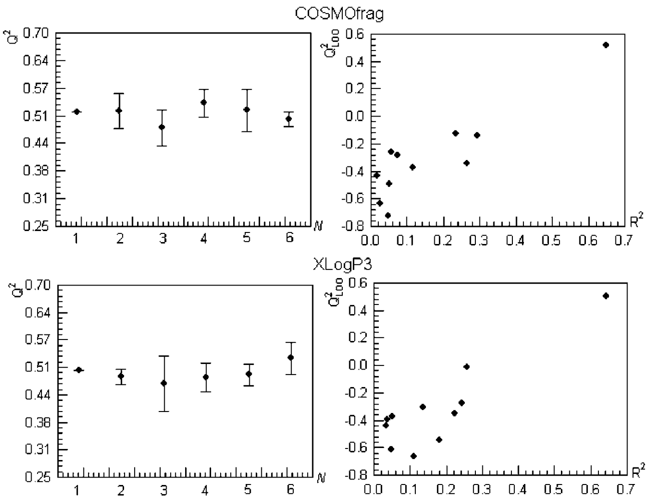

Figure 5. Plots for LNO cross-validation (left) and **y**-randomization test (right) for the models COSMOfrag and XLog*P*3 of TS3.

present the best *r*'s explain and use larger amount of original information, and have the best results for the external validation.

On the other hand, the models with the descriptors KOWWIN, ChemOffLog*P* and, specially, CLog*P*, had the poorest statistics. In the external prediction, R^2_{pred} values (less than 0.1) were unacceptable and show no correlation between the experimental and predicted activities. The three models also failed in the LNO validation and **y**-randomization tests (Fig. 6). The results from **y**-randomization test and LNO cross-validation of TS2 are available in the Supplementary Material, Tables S17 and S18.

3.5 Overview of the Results

Having in mind the questions raised initially, it is rather clear that the use of any algorithm to calculate Log*P*'s without an a priori selection or comparison among them, can lead to poor results in a QSAR study. Distinct algorithms can contribute with different amounts and types of information encoded in Log*P*'s, leading to models with reasonable statistical differences, as occurred with TS3.

Although the selected Log*P* descriptor for the new study (TS3) has the highest *r* (Table 4), this is not sufficient to generate the best multivariate regression model. This fact becomes clear when analyzing the Huuskonen's data [27], in which the algorithm used by the author yielded the second worst *r*. The same can be said about the Bansal and co-workers' data [26].

The performance of the models in all validations carried out also have shown to be important in the QSAR studies. In the case of TS1, the external validation aided to choose the best model. For TS2, despite that all the models appeared to be statistically equivalent, two of were based on chance correlation. Thus, to select the most appropriate algorithm for Log*P* calculation for each case in the present work, comparison between the QSAR models fully validated had to be carried out.

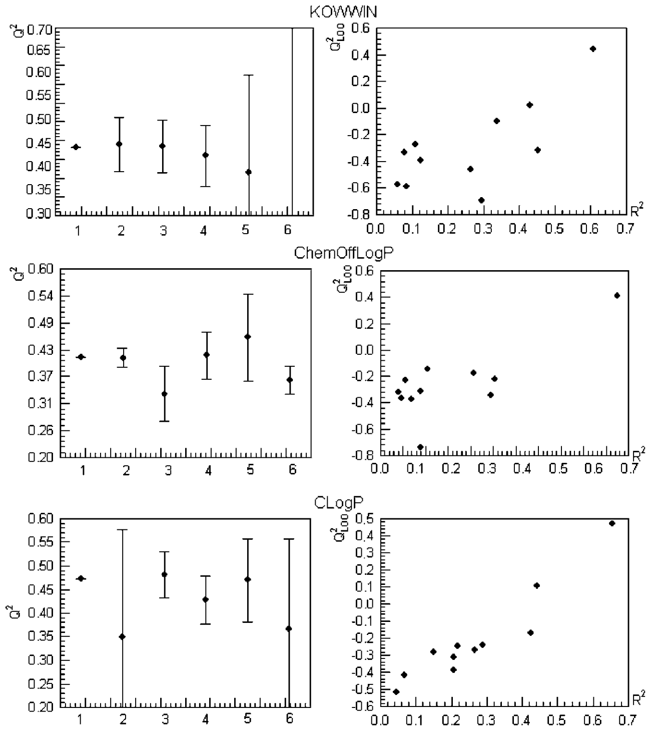

Figure 6. Plots for LNO cross-validation (left) and **y**-randomization test (right) for the models KOWWIN, ChemOffLog*P* and CLog*P* of TS3.

Finally, it is possible to observe that there is no unique algorithm leading always to the highest quality QSAR models, as comes out clearly from the present analyses.

In a previous work from our group [25], the problem of the most relevant lipophilicity descriptor(s) in 3 regression models for β-lactam inhibitors of 3 strains of *Salmonella thypimurium* was posed, and solved by exploratory and PLS analyses. β-Lactam antibiotics belong to a specific class of organic compounds for which lipophilicity is an essential determinant of variations in antibacterial activity. The calculated lipophilicity descriptors were not of pure lipophilic nature, but included various steric and electronic features, because of which they behaved as general descriptors during the variable selection (more than one lipophilicity descriptor was selected). These trends were observed for all lipophilicity descriptors (7 Log*P*'s and 2 non-Log*P*'s), meaning that the problem of the choice of the most relevant Log*P* may be extended to other types of lipophilicity descriptors.

In another QSAR approach [44] the same β-lactam inhibitors of *S. thypimurium* were described by another set of descriptors, denominated a priori, mainly topologically derived, and some of them were considered as amphiphilicity descriptors. The two studies about the β-lactams have shown rather clearly that the calculated Log*P*'s could be replaced by steric, electronic, topological and combined descriptors. Such a situation indicates that distinct algorithms for Log*P* calculations in QSAR studies may result in descriptors of rather different contents of liphophilic nature. This is probably the reason why PLS models containing different Log*P*'s in the present work can be distinguished in terms of statistical parameters and model validations, especially in the case of TS3.

Lipophilicity is a property which is always important for biological activity of a drug because it is a measure of drug's interaction with any kind of media (hydrophobic, amphyphilic, hydrophilic, lipophilic, etc.). However, this does not imply always that variations in lipophilicity for a set of drugs are important for the variations in the respective biological activity. In the absence of experimental Log*P*, it is recommended that the evaluation of lipophilicity's role in drug action is carried out in the following steps: 1) calculation of Log*P*'s (and eventually other lipophilicity parameters) by diverse algorithms; 2) inclusion of the obtained descriptors in the total descriptors pool; 3) variable selection, construction of the final regression model and its complete validation.

In fact, the problem of the most relevant lipophilicity descriptor in a QSAR study may be extended to other types of molecular descriptors which are sensitive to calculation procedures performed: atomic charges, dipole moment and its components, polarizability, hyperpolarizabilities and their components, and so on.

4 Conclusions

The results strengthen the hypothesis that, when the experimental values of Log*P* are not available, the choice of an algorithm for calculation of Log*P*, from chemical structures, may influence the final results of a QSA(P)R study. Among the tested algorithms, two of the most suitable to relate the lipophilicity of each training set with the biological activities were CSLog*P* for TS1 and CLog*P* for TS2. Both algorithms are commercial and, in this case, a good alternative would be the use of the freeware algorithm MLog*P* for TS1 (as in the original work) and KOWWIN for TS2.

It is noteworthy that the results presented in this work have no intention to delegate more or less relevance to the tested algorithms, to consider some as appropriate or not for any QSA(P)R study or to quarrel the results from other research groups. Different training sets and activities (or properties) have its own characteristics, and the same can be said regarding to the Log*P* algorithms. Because of this fact, experimental values, when available, should be always the first choice to obtain more realistic models.

For QSAR studies where Log*P* is important to describe the drug mechanism of action and for which no experimental data are available, it is highly recommended to proceed with the procedure suggested in this work, taking into account the availability of freeware softwares.

5 Acknowledgements

EBM thanks the Universidade Estadual do Oeste do Paraná (http://www.unioeste.br) for supporting in the doctoral thesis, and the Institute of Chemistry of the University of Campinas (http://www.iqm.unicamp.br). MMCF acknowledges FAPESP for research funding (2004/04686-5).

6 References

[1] H. Kubinyi, *QSAR: Hansch Analisys and Related Approaches*, Wiley-VCH, Weinheim **1993**, pp. 21 – 56.

[2] A. T. Florence, D. Attwood, *Princípios Físico-Químicos em Farmácia,* Edusp, São Paulo **2003**, pp. 219 – 278.

[3] W. M. Meylen, P. H. Howard, *J. Pharm. Sci.* **1995**, *84*, 83 – 92.

[4] A. Breindl, B. Beck, T. Clark, R. C. Glen, *J. Mol. Model.* **1997**, *3*, 142 – 155.

[5] C. Hattotuwagama, D. R. Flower, *Bioinformation* **2006**, *1*, 257 – 259.

[6] M. Medic-Saric, A. Mormar, J. Jasprica, *Acta Pharm.* **2004**, *54*, 91 – 101.

[7] G. E. Kellog, D. J. Abraham, *Eur. J. Med. Chem.* **2000**, *35*, 651 – 661.

[8] Y. Sakwatani, K. Kasai, Y. Noguchi, J. Yamada, *QSAR Comb. Sci.* **2007**, *26*, 109 – 116.

[9] F. A. L. Ribeiro, M. M. C. Ferreira, *J. Mol. Struct.-Theochem.* **2003**, *663*, 109 – 126.

[10] G. L. Patrick, *An Introduction to Medicinal Chemistry*, Oxford, New York, **2001**, pp. 128 – 153.

[11] *Medicinal Chemistry: Principles and Practice* (Ed: F. D. King), RSC, Cambridge **2002**, pp. 195 – 214.

[12] S. A. Teijeiro, G. N. Moroni, M. I. Motina, M. C. Briñón, *J. Liq. Chrom. Rel. Technol.* **2000**, *23*, 855 – 872.

[13] Y. Zhao, J. Jona, D. T. Chow, H. Rong, D. Semin, X. Xia, R. Zanon, C. Spancake, E. Maliski, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1548 – 1555.

[14] J. E. A. Conner, A. Curdeef, K. J. Box, *American Laboratory* **1995**, *27*, 36C – 36C.

[15] http://www.biobyte.com/bb/prod/cqsar.html (acessed August 31, 2008).

[16] http://www.syrres.com/esc/physprop.htm (acessed August 31, 2008).

[17] T. Fujita, J. Iwasa, C. Hansch, *J. Am. Chem. Soc.* **1964**, *86*, 5175 – 5180.

[18] R. Mannhold, H. van Waterbeend, *J. Comp.-Aided Mol. Design* **2001**, *15*, 337 – 354.

[19] G. Thomas, *Química Medicinal: uma Introdução*, Guanabara Koogan, Rio de Janeiro **2003**, pp. 23 – 71.

[20] J. V. Tetko, *Mini Rev. Med. Chem.* **2003**, *3*, 809 – 820.

[21] R. Mannhold, *Mini Rev. Med. Chem.* **2005**, *5*, 197 – 205.

[22] E. Benfenati, G. Gini, N. Piclin, A. Roncaglioni, M. R. Vari, *Chemosphere* **2003**, *53*, 1155 – 1164.

[23] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *J. Pharm. Sci.* **2009**, *98*, 861 – 893.

[24] M. Karthikeyan, S. Krishnan, A. K. Pandey, A. Bender, A. Tropsha, *J. Chem. Inf. Model.* **2008**, *48*, 691 – 703.

[25] M. M. C. Ferreira, R. Kiralj, *J. Chemometr.* **2004**, *18*, 242–252.
[26] R. Bansal, C. Karthikeyan, N. S. H. N. Moorthy, P. Trivedi, *Arkivoc* **2007**, *15*, 66–81.
[27] J. Huuskonen, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 425–429.
[28] A. Petrocchi, U. Koch, V. G. Matassa, B. Pacini, K. A. Stillmockb, V. Summa, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 350–353.
[29] M. L. Barreca, L. De Luca, S. Ferro, A. Rao, A. Monforte, A. Chimirri, *Arkivoc* **2006**, *7*, 224–244.
[30] *Matlab 7*, MathWorks Inc., Natik, USA **2006**.
[31] M. M. C. Ferreira, *J. Braz. Chem. Soc.* **2002**, *13*, 742–753.
[32] *Pirouette 4*, Infometrix Inc., Woodinville, USA **2007**.
[33] *Chemometric Methods in Molecular Design*, (Ed: H. van de Waterbeemd), Wiley-VCH, Weinheim **1995**, pp. 15–38.
[34] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Chem.* **2003**, *22*, 69–77.
[35] A. Golbraikh, A. Tropsha, *J. Mol. Grap. Modell*. **2002**, *20*, 269–276.
[36] P. Gramática, *QSAR Comb. Chem.* **2007**, *26*, 694–701.
[37] A. O. Aptula, N. G. Jeliazkova, T. W. Schultz, M. T. D. Cronin, *QSAR Comb. Chem.* **2005**, *24*, 385–396.
[38] C. Rücker, G. Rücker, M. Meringer, *J. Chem. Inf. Model.* **2007**, *47*, 2345–2357.
[39] L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramática, *Environ. Health Perspect.* **2003**, *111*, 1361–1375.
[40] A. C. Gaudio, E. Zandonade, *Quim. Nova* **2001**, *24*, 658–671.
[41] A. A. M. Chasin, E. S. Nascimento, L. M. Ribeiro-Neto, M. E. P. B. Siqueira, M. H. Andraus, M. C. Salvadori, N. A. G. Fernícola, R. Gorni, S. Salcedo, *Rev. Bras. Toxicol.* **1998**, *11*, 1–6.
[42] DataFit 9, Oakdale Engineering, Oakdale, USA **2008**.
[43] G. Melagraki, A. Afantitis, H. Sarimveis, P. A. Koutentis, J. Markopolus, O. Igglessi-Markopoulou, *J. Comput. Aided. Mol. Des.* **2007**, *21*, 251–267.
[44] R. Kiralj, M. M. C. Ferreira, *Croat. Chem. Acta* **2008**, *81*, 579–592.